



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12N 15/12, 15/54, 9/12 // C12Q 1/68, C12N 15/11	A1	(11) International Publication Number: WO 93/15201 (43) International Publication Date: 5 August 1993 (05.08.93)
(21) International Application Number: PCT/US93/00586 (22) International Filing Date: 22 January 1993 (22.01.93) (30) Priority data: 826,935 22 January 1992 (22.01.92) US (60) Parent Application or Grant (63) Related by Continuation US 826,935 (CIP) Filed on 22 January 1992 (22.01.92) (71) Applicant (for all designated States except US): NEW ENGLAND DEACONESS HOSPITAL [US/US]; 185 Pilgrim Road, Boston, MA 02215 (US).	(72) Inventors; and (75) Inventors/Applicants (for US only) : AVRAHAM, Hava [IL/US]; 50 Radmor Road, Brighton, MA 02135 (US). GROOPMAN, Jerome [US/US]; 79 Druce Street, Brookline, MA 02146 (US). COWLEY, Sally [GB/GB]; 28 Avonmore Road, London W14 8RS (GB). SCADDEN, David [US/US]; 62 Lexington Street, Weston, MA 02193 (US). (74) Agents: GRANAHAH, Patricia et al.; Hamilton, Brook, Smith & Reynolds, Two Militia Drive, Lexington, MA 02173 (US). (81) Designated States: AU, CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>	
(54) Title: NOVEL PROTEIN TYROSINE KINASES		
(57) Abstract <p>The identification and isolation of novel protein tyrosine kinase genes present on human megakaryocytic and lymphocytic cells, the proteins encoded by these genes, antibodies specific for the encoded proteins, RNA nucleic acid sequences which hybridize to the genes and methods of use therefor.</p>		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	MR	Mauritania
AU	Australia	GA	Gabon	MW	Malawi
BB	Barbados	GB	United Kingdom	NL	Netherlands
BE	Belgium	GN	Guinea	NO	Norway
BF	Burkina Faso	GR	Greece	NZ	New Zealand
BG	Bulgaria	HU	Hungary	PL	Poland
BJ	Benin	IE	Ireland	PT	Portugal
BR	Brazil	IT	Italy	RO	Romania
CA	Canada	JP	Japan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SK	Slovak Republic
CI	Côte d'Ivoire	LJ	Liechtenstein	SN	Senegal
CM	Cameroon	LK	Sri Lanka	SU	Soviet Union
CS	Czechoslovakia	LJ	Luxembourg	TD	Chad
CZ	Czech Republic	MC	Monaco	TG	Togo
DE	Germany	MG	Madagascar	UA	Ukraine
DK	Denmark	ML	Mali	US	United States of America
ES	Spain	MN	Mongolia	VN	Viet Nam
FI	Finland				

NOVEL PROTEIN TYROSINE KINASESDescriptionBackground of the Invention

Transduction of signals that regulate cell growth and differentiation is regulated in part by phosphorylation of various cellular proteins. Protein tyrosine kinases are enzymes that catalyze this process. Moreover, many act as growth factor receptors.

Summary of the Invention

10 The present invention relates to novel protein tyrosine kinase genes present in human megakaryocytic and lymphocytic cells, the proteins encoded by these genes, antibodies specific for the encoded proteins, RNA nucleic acid sequences which hybridize to the genes and methods of use therefor.

The genes isolated as described herein are referred to, collectively, as protein tyrosine kinase (pTK) genes. The nucleic acid sequences of these genes, isolated as discussed herein, show significant homology with previously identified protein tyrosine kinases containing extracellular domains which function as growth factor receptors. The pTK genes have been shown to be present in both megakaryocytic and lymphocytic cells.

The pTK genes of the present invention show significant sequence homology with members of the c-kit subgroup of growth factor receptors with protein tyrosine kinase activity. The c-kit subgroup of receptor tyrosine kinases catalyze the phosphorylation of exogenous substrates, as well as tyrosine residues within their own polypeptide chains. (Ullrich, A. and Schlessinger, J., Cell, 61:203 (1990)). Members of the c-kit subgroup include FLT/FLK (Fetal Liver Kinase), FGF (Fibroblast

-2-

Growth Factor Receptor) and NGF (Nerve Growth Factor Receptor).

In particular, fourteen pTK genes have been identified. Two pTK genes, referred to as SAL-S1 and SAL-D4 (also referred to as megakaryocyte derived FGF-like receptor tyrosine kinase) were identified in megakaryocytic cells. Five pTK genes, referred to as LpTKs, were identified in lymphocytic cells and have been shown to be present in megakaryocytes as well. One pTK gene, referred to as HpTKs, was identified in human hepatoma cells. Six pTK genes, referred to as bpTK genes, found in human brain tissue.

SAL-S1 is related to the FLT/FLK family of pTKs. SAL-D4 is related to the FGF receptor family of pTKs, and one LpTK (LpTK 3) is related to the NGF receptor family of pTKs.

The pTK genes, which are the subject of the present invention, were identified using two sets of degenerative oligonucleotide primers: a first set which amplifies all pTK DNA segments (SEQ ID NOS:1-2), and a second set which amplifies highly conserved sequences present in the catalytic domain of the c-kit subgroup of pTKs (SEQ ID NOS:3-4). The pTK genes identified in this manner are described below.

SAL-S1 is expressed in several megakaryocytic cell lines, but not in erythroid cell lines. The nucleotide sequence of SAL-S1 was obtained, revealing a sequence containing 158 base pairs. (SEQ ID NO:5). This isolated DNA fragment encoded an amino acid sequence (SEQ ID NO:6) which exhibited significant sequence homology with known protein tyrosine kinases of the FLT/FLK family. The full

-3-

length gene sequence (SEQ ID NO: 17) contains 6827 b.p. and the deduced amino acid sequence (SEQ ID NO: 18) contains 349 residues.

SAL-D4, also expressed in megakaryocytic cells, is a DNA fragment containing the nucleotide sequence of 141 base pairs. (SEQ ID NO:7). This isolated DNA fragment encoded an amino acid sequence (SEQ ID NO:8) which exhibited significant sequence homology with known protein tyrosine kinases of the FGF receptor family.

10 The LpTKs, including LpTK 2, LpTK 3, LpTK 4, and LpTK 13 and LpTK 25, are expressed in lymphocytic cells, as well as megakaryocytic cells. The nucleotide sequence (151 base pairs) of the LpTK 3 gene was obtained (SEQ ID NO:11), and exhibited significant homology with known
15 protein tyrosine kinases of the NGF receptor family. The nucleotide sequences of the LpTK 2, LpTK 4, and LpTK 13 genes contained 149 base pairs (SEQ ID NO:9), 137 base pairs (SEQ ID NO:13), and 211 base pairs (SEQ ID NO:15) respectively. LpTK 25 has a nucleotide sequence of 3120
20 b.p. (SEQ ID NO: 22). A full length gene sequence has been obtained for LpTK 2 (SEQ ID NO: 19) which contains 7606 b.p. Additional sequencing of LpTK 4 revealed a sequence of 404 b.p. (SEQ ID NO: 21).

The HpTK 5 gene, expressed in human hepatoma cells,
25 has a nucleotide sequence of 3120 b.p. (SEQ ID NO: 22). Nucleotide sequences of the bpTK's, including bpTK 1, bpTK 2, bpTK 3, bpTK 4, bpTK 5 and bpTK 7 are expressed in human brain tissue encode proteins having the amino acid sequences of SEQ ID NOS: 25-30 respectively.

30 Thus the present invention includes DNA isolated from a human megakaryocytic cell line, which hybridizes to a

-4-

DNA fragment which hybridizes to DNA encoding an amino acid sequence which is highly conserved in the catalytic domain of protein tyrosine kinases of the c-kit subgroup.

The present invention also includes the proteins
5 encoded by the pTK genes identified as described herein, which exhibit significant sequence homology with members of the c-kit subgroup of pTKs (i.e. FLT/FLK (SAL-S1), FGF receptor (SAL-D4) or NGF receptor (LpTKS)) as well as the proteins encoded by HpTK 5 and the bpTKs. The present
10 invention also includes SAL-S1, SAL-D4, and LpTK, HpTK and bpTK homologues or equivalents (i.e., proteins which have amino acid sequences substantially similar, but not identical, to that of SAL-S1, SAL-D4, the LpTKs HpTK and the bpTKs, which exhibit tyrosine kinase activity.) This
15 invention further includes peptides (SAL-S1, SAL-D4, LpTK, HpTK and bpTK fragments) which retain tyrosine kinase activity, yet are less than the entire SAL-S1, SAL-D4, LpTK, HpTK and bpTK sequences), monoclonal and polyclonal antibodies specific for SAL-S1, SAL-D4, the LpTKs, HpTK
20 and the bpTKs, and uses for the SAL-S1, SAL-D4, the LpTK, HpTK and the bpTK nucleic acid sequences and SAL-S1, SAL-D4, LpTK, HpTK and bpTK equivalents.

The present invention further includes nucleic acid sequences which hybridize with DNA or RNA encoding the
25 proteins described herein, which exhibit significant sequence homology with the FLT/FLK, FGF receptor or NGF receptor family of protein tyrosine kinases contained within the c-kit subgroup. Such nucleic acid sequences are useful as probes to identify pTK genes in other
30 vertebrates, particularly mammals, and in other cell types. They can also be used as anti-sense

-5-

oligonucleotides to inhibit protein tyrosine kinase activity, both in vitro and in vivo.

The SAL-S1, SAL-D4, LpTK, HpTK and bpTK, tyrosine kinases of the present invention can be used as target
5 proteins in conjunction with the development of drugs and therapeutics to modulate cell growth, differentiation and other metabolic functions. The SAL-S1, SAL-D4, LpTK, HpTK or bpTK proteins can be used as agonists or antagonists to other tyrosine kinases. The SAL-S1, SAL-D4, LpTK, HpTK or
10 bpTK tyrosine kinases can also be instrumental in the modulation of megakaryocyte and/or platelet adhesion interactions.

In addition, the SAL-S1, SAL-D4, LpTK, HpTK and bpTK tyrosine kinases can be used in screening assays to detect
15 cellular growth and/or differentiation factors. Using standard laboratory techniques, the ligands of the pTKs of the present invention can be identified. Once identified, assays can be designed to detect these ligands present endogenously, within cells, as well as exogenously, in
20 extra cellular fluids. Assays can also be designed as diagnostic aids to detect these ligands in body fluids such as blood and urine.

Brief Description of the Drawings

Figure 1 depicts the nucleotide sequence of SAL-S1
25 (SEQ ID NO: 5) and the deduced amino acid sequence (SEQ ID NO:6).

Figure 2 depicts the nucleotide sequence of SAL-D4 (SEQ ID NO:7) and its deduced amino acid sequence (SEQ ID NO:8).

-6-

Figure 3A depicts the nucleotide sequence (SEQ ID NO:9) and its deduced amino acid sequence (SEQ ID NO:10) for LpTK 2.

Figure 3B depicts the nucleotide sequence (SEQ ID NO:11) and its deduced amino acid sequence (SEQ ID NO:12) for LpTK 3.

Figure 3C depicts the nucleotide sequence (SEQ ID NO:13) and its deduced amino acid sequence (SEQ ID NO:14) for LpTK 4.

10 Figure 3D depicts the nucleotide sequence (SEQ ID NO:15) and its deduced amino acid sequence (SEQ ID NO:16) for the LpTK 13.

Figure 4A-4J depicts the full-length nucleotide sequence (SEQ ID NO: 17) and its deduced amino acid sequence (SEQ ID NO: 18) for SAL-S1.

Figure 5A-5J depicts the full length nucleotide sequence (SEQ ID NO: 19) and the deduced amino acid sequence (SEQ ID NO: 20) for LpTK2.

20 Figure 6 depicts the partial nucleotide sequence (SEQ ID NO: 21) for LpTK4.

Figure 7A-7D depicts the full length nucleotide sequence (SEQ ID NO: 22) for LpTK25.

25 Figure 8A-8F depicts the full length nucleotide sequence (SEQ ID NO: 23) and the deduced amino acid sequence (SEQ ID NO: 24) for HpTK5.

Figure 9 depicts the amino acid sequence (SEQ ID NO: 25) of bpTK1.

Figure 10 depicts the amino acid sequence (SEQ ID NO: 26) of bpTK2.

30 Figure 11 depicts the amino acid sequence (SEQ ID NO: 27) of bpTK3.

-7-

Figure 12 depicts the amino acid sequence (SEQ ID NO: 28) of bpTK4.

Figure 13 depicts the amino acid sequence (SEQ ID NO: 29) of bpTK5.

5 Figure 14 depicts the amino acid sequence (SEQ ID NO: 30) of bpTK7.

Detailed Description of the Invention

Novel protein tyrosine kinase genes have been identified, their nucleic acid sequences determined, and
10 the amino acid sequences of the encoded proteins deduced. The genes isolated as described herein are referred to, collectively, as protein tyrosine kinase (pTK) genes. The nucleic acid sequences of these genes, isolated as discussed herein, show significant homology to
15 with previously identified protein tyrosine kinases containing extracellular domains which function as growth factor receptors. These genes have been shown to be present in both megakaryocytic and lymphocytic cells.

To facilitate the isolation and identification of
20 these novel pTKs, two sets of DNA probes were used, as described in the Exemplification. The first set consisted of two degenerative oligonucleotide sequences, pTK 1 (SEQ ID NO:1) and pTK 2 (SEQ ID NO:2) (Matthews, W. Cell 65: 1143 (1991; Wilks, A. F. Proc. Natl. Acad. Sci. USA
25 86:1603 (1989)). These sequences were used as primers in a polymerase chain reaction to amplify tyrosine kinase DNA segments. (Mullis, K. et al., Cold Spring Harbor Symp. Advan. Biol. 51:263 (1986).

The second set consisted of two oligonucleotide
30 sequences, pTK 3 (SEQ ID NO:3) and pTKKW (SEQ ID NO:4)

-8-

designed to amplify the nucleic acid sequence which encodes the highly conserved regions of the catalytic domains of the c-kit family of protein tyrosine kinases. These sequences were used as primers in the polymerase chain reaction in a second round of DNA amplification. Using this two-step amplification procedure, DNA fragments which hybridized to these pTK primers were identified, isolated and subsequently sequenced.

In particular, fourteen pTK genes exhibiting significant homology with the c-kit subgroup of protein tyrosine kinases have been identified. Two pTK genes, referred to as SAL-S1 and SAL-D4 (also referred to as megakaryocyte derived FGF-like receptor) were identified in several megakaryocytic cell lines, including CMK 11-5, DAMI, UT-7 and UT-7 grown in erythropoietin, but not in the erythroid cell lines HEL, PMA stimulated HEL cells, or K562. Five pTK genes, referred to as LpTKs, were identified in lymphocytic, as well as in megakaryocytic cells. One pTK gene, referred to as HpTK5 was identified in human hepatoma cells and six genes, referred to as bpTKs, were identified in human brain tissue.

SAL-S1 (SEQ ID NO:6 and 18) encoded by the nucleic acid sequence SEQ ID NOS:5 and 17, exhibits significant homology with the FLT/FLK family of pTKs. SAL-D4 (SEQ ID NO:8) encoded by SEQ ID NO:7, is related to the FGF receptor family of pTKs, and one LpTK (LpTK 3 (SEQ ID NO:12) encoded by the SEQ ID NO:11, is related to the NGF receptor family of pTKs. The remaining LpTKs, LpTK2 (SEQ ID NO:10) encoded by SEQ ID NO:9; LpTK4 (SEQ ID NO:14) encoded by SEQ ID NO:13; LpTK13 (SEQ ID NO:16) encoded by SEQ ID NO:15 LpTK25 encoded by SEQ ID NO: 22, also exhibit

-9-

sequence homology with known protein tyrosine kinases
(Data not shown).

HpTK5 (SEQ ID NO: 24) encoded by SEQ ID NO: 23 and
the bpTKs 1, 2, 3, 4, 5 and 7 (SEQ ID NOS: 25-30
5 respectively, also exhibit sequence homology with known
protein tyrosine kinases.

Thus, as described above, DNA which hybridize with
DNA encoding amino acid sequences present in the catalytic
domain of a protein tyrosine kinase of the c-kit subgroup
10 of protein kinases have been isolated and sequenced.
These isolated DNA sequences, collectively referred to as
pTKs genes, (and their deduced amino acid sequences) have
been shown to exhibit significant sequence homology with
known members of receptor tyrosine kinase families.

15 Once isolated, these DNA fragments can be amplified
using known standard techniques such as PCR. These
amplified fragments can then be cloned into appropriate
cloning vectors and their DNA sequences determined.

These DNA sequences can be excised from the cloning
20 vectors, labeled with a radiolabeled nucleotide such as
³²P and used to screen appropriate cDNA libraries to
obtain the full-length cDNA clone.

The pTk genes as described above have been isolated
from the source in which they occur naturally, i.e.
25 megakaryocyte and lymphocytic cells. The present invention
is intended to include pTk genes produced using genetic
engineering techniques, such as recombinant technology, as
well as pTk genes that are synthesized chemically.

The deduced amino acid sequences of the pTK genes
30 include amino acid sequences which encode peptides
exhibiting significant homology with the catalytic domain

-10-

of protein tyrosine kinases of the c-kit subgroup of tyrosine kinases. These proteins, encoded by the pTk genes, can include sequences in which functionally equivalent amino acid residues are substituted for
5 residues within the sequence, resulting in a silent change, that is a change not detected phenotypically. For example, one or more amino acid residues within the sequence can be substituted by another amino acid of a similar polarity which acts as a functional equivalent,
10 resulting in a silent substitution.

In addition, the protein structure can be modified by deletions, additions, inversion, insertions or substitutions of one or more amino acid residues in the sequence which do not substantially detract from the
15 desired functional tyrosine kinases properties of the peptide.

Modified pTKs of the present invention, with receptor tyrosine kinase activity can be made using recombinant DNA techniques, such as excising it from a vector containing a
20 cDNA encoding such a protein, or by synthesizing DNA encoding the desired protein mechanically and/or chemically using known techniques.

An alternate approach to producing the pTKs of the present invention is to use peptide synthesis to make a
25 peptide or polypeptide having the amino acid sequence of such a protein. The peptides or modified equivalents thereof, can be synthesized directly by standard solid or liquid phase chemistries for peptide synthesis.

Preferably, the pTKs of the present invention will be
30 produced by inserting DNA encoding the proteins into an appropriate vector/host system where it will be expressed.

-11-

The DNA sequences can be obtained from sources in which they occur naturally, can be chemically synthesized or can be produced using standard recombinant technology.

This invention also pertains to an expression vector
5 comprising a pTK gene of the present invention, encoding for a protein which exhibits receptor tyrosine kinase activity.

The pTK genes of the present invention can be used for a number of diagnostic and therapeutic purposes. For
10 example, the nucleic acid sequences of the pTK genes can be used as probes to identify other protein tyrosine kinases present in other cell types, including eukaryotic and prokaryotic cell types.

The nucleic acid sequences can be used to design
15 drugs that directly inhibit the kinase activity of protein tyrosine kinases, or to design peptides that bind to the catalytic domain of tyrosine kinases, thus inhibiting their activity. These sequences can also be used to design anti-sense nucleotides that can also inhibit, or
20 destroy, tyrosine kinase activity. Such inhibition of tyrosine kinase activity would be desirable in pathological states where decreased cellular proliferation would be beneficial, such as leukemias or other malignancies.

25 The nucleic acid sequences can also be used to design drugs, peptides or anti-sense nucleotides as above, but with enhancing, rather than inhibitory effects, on tyrosine kinases. Such enhanced tyrosine kinase activity would result in increasing the phosphorylation of
30 substrates (exogenous, as well as endogenous tyrosine residues). Enhanced effects would be desirable in states

-12-

where increased cellular proliferation would be beneficial, such as anemias, bleeding disorders and during surgical procedures.

The pTK genes of the present invention can also be
5 used to obtain soluble fragments of receptor tyrosine kinases, capable of binding their respective ligands (i.e. fibroblast growth factor).

pTK genes encoding soluble receptor tyrosine kinase fragments can be produced using recombinant DNA techniques
10 or synthetically. In either case, the DNA obtained encodes a soluble pTK fragment which lacks a substantial portion of the hydrophobic transmembrane region to permit solubilization of the fragment.

These soluble pTK protein fragments can be introduced
15 exogenously to act as competitors with the endogenous, membrane bound pTK for their respective ligands, thus inhibiting tyrosine kinase activity. Alternately, a modified soluble pTK protein fragment can be introduced which binds the ligand but does not activate kinase
20 activity.

These soluble pTK protein fragments can also be used in binding assays to detect ligands such as growth and differentiation factors. Once these ligands are identified, they may be altered or modified to inhibit or
25 enhance kinase activity. For example, the ligands may be modified or attached to substances that are toxic to the cell, such a ricin, thus destroying the target cell. The substance may be a super-activating substance which, after binding to the pTK, may substantially increase the kinase
30 activity, or activate other growth factors.

-13-

pTk genes of the present invention would also be useful to develop diagnostic tools for in vitro screening assays for ligands such as growth factors or differentiation factors that inhibit or enhance kinase activity. The proteins encoded by the pTK genes can also be used in such assays, or as immunogens to produce monoclonal or polyclonal antibodies to be used in such assays.

Such antibodies can also be used in methods of treating conditions in which an individual would benefit therapeutically if protein tyrosine kinase activity could be modified, such as increasing platelet production in bleeding disorders.

The present invention will now be illustrated by the following Exemplification, which is not intended to be limiting in any way.

Exemplification: The Identification and Isolation of the pTK Genes

To facilitate the isolation and identification of these novel pTK genes, two sets of DNA probes were used. (See Table).

The first set consisted of two degenerative oligonucleotide sequences, pTK 1 (SEQ ID NO:1) and pTK 2 (SEQ ID NO:2). These sequences were used as polymerase chain reaction (PCR) primers, using standard PCR techniques, to amplify tyrosine kinase DNA segments.

The second set consisted of two oligonucleotide sequences, pTK 3 (SEQ ID NO:3) and pTKKW (SEQ ID NO:4) selected from the highly conserved regions of the catalytic domains of the c-kit subgroup of protein

-14-

tyrosine kinases. These sequences were also used as polymerase chain reaction primers in a second round of DNA amplification. Using this two-step amplification procedure, DNA fragments which hybridized to these pTK
5 primers were identified, isolated and subsequently sequenced using known laboratory techniques.

TABLEFirst Round of Amplification

PTK1

10 CGGATCCACAGNGACCT

PTK2

GGAATTCCAAAGGACCAGACGTC

Second Round of Amplification

PTK3 (kit family specific)

15 CGGATCCATCCACAGAGATGT

PTKKW (kit family specific)

GGAATTCCTTCAGGAGCCATCCACTT

Equivalents

Those skilled in the art will recognize, or be able
20 to ascertain using no more than routine experimentation,
many equivalents to the specific embodiments of the
invention described herein. Such equivalents are intended
to be encompassed by the following claims.

-15-

CLAIMS

The invention claimed is:

1. Isolated DNA of human megakaryocytic origin which hybridizes to a DNA fragment which hybridizes to DNA
5 encoding an amino acid sequence present in the catalytic domain of a protein tyrosine kinase of the c-kit subgroup of protein tyrosine kinases.
2. Isolated DNA of Claim 1 having a nucleotide sequence selected from the group of nucleotide sequences
10 consisting of:
 - a) SAL-S1 (SEQ ID NOS:5 and 7);
 - b) SAL-D4 (SEQ ID NO:7);
 - c) LpTK 2 (SEQ ID NOS:9 and 19);
 - d) LpTK 3 (SEQ ID NO:11);
 - 15 e) LpTK 4 (SEQ ID NOS:13 and 21);
 - f) LpTK 13 (SEQ ID NO:15);
 - g) LpTK 25 (SEQ ID NO: 22 and
 - h) HpTK 5 (SEQ ID NO: 23).
3. Isolated DNA of Claim 1 which encodes an amino acid
20 sequence selected from the group consisting of:
 - a) SAL-S1 (SEQ ID NOS:6 and 18);
 - b) SAL-D4 (SEQ ID NO:8);
 - c) LpTK 2 (SEQ ID NOS:10 and 20);
 - d) LpTK 3 (SEQ ID NO:12);
 - 25 e) LpTK 4 (SEQ ID NO:14);
 - f) LpTK 13 (SEQ ID NO:16);
 - g) HpTK 5 (SEQ ID NO:24);

-16-

- h) bpTK 1 (SEQ ID NO:25);
- i) bpTK 2 (SEQ ID NO:26);
- j) bpTK 3 (SEQ ID NO:27);
- k) bpTK 4 (SEQ ID NO:28);
- 5 l) bpTK 5 (SEQ ID NO:29); and
- m) bpTK 7 (SEQ ID NO:30).

4. Isolated DNA of human megakaryocytic origin which comprises a DNA fragment whose sequence encodes an amino acid sequence present in the catalytic domain
10 of a protein tyrosine kinase of the c-kit subgroup of protein tyrosine kinases.

5. Isolated DNA of Claim 4 which encodes an amino acid sequence selected from the group consisting of:

- a) SAL-S1 (SEQ ID NO:6);
- 15 b) SAL-D4 (SEQ ID NO:8);
- c) LpTK 2 (SEQ ID NO:10);
- d) LpTK 3 (SEQ ID NO:12);
- e) LpTk 4 (SEQ ID NO:14); and
- f) LpTK 13 (SEQ ID NO:16).
- 20 g) HpTK 5 (SEQ ID NO:24);
- h) bpTK 1 (SEQ ID NO:25);
- i) bpTK 2 (SEQ ID NO:26);
- j) bpTK 3 (SEQ ID NO:27);
- k) bpTK 4 (SEQ ID NO:28);
- 25 l) bpTK 5 (SEQ ID NO:29); and
- m) bpTK 7 (SEQ ID NO:30).

-17-

6. A homogeneous protein of human megakaryocytic origin which includes an amino acid sequence exhibiting sequence homology with the catalytic domain of tyrosine kinases of the c-kit family.
- 5 7. A homogeneous protein of Claim 6 in which the amino acid sequence is selected from the group consisting of:
 - a) SAL-S1 (SEQ ID NO:6);
 - b) SAL-D4 (SEQ ID NO:8);
 - 10 c) LpTK 2 (SEQ ID NO:10);
 - d) LpTK 3 (SEQ ID NO:12);
 - e) LpTk 4 (SEQ ID NO:14); and
 - f) LpTK 13 (SEQ ID NO:16).
 - g) HpTK 5 (SEQ ID NO:24);
 - 15 h) bpTK 1 (SEQ ID NO:25);
 - i) bpTK 2 (SEQ ID NO:26);
 - j) bpTK 3 (SEQ ID NO:27);
 - k) bpTK 4 (SEQ ID NO:28);
 - l) bpTK 5 (SEQ ID NO:29); and
 - 20 m) bpTK 7 (SEQ ID NO:30).
8. A protein of human megakaryocytic origin which exhibits significant sequence homology with the FLT/FLK family of protein tyrosine kinases.
9. A protein of Claim 8 encoded by the nucleotide
25 sequence (SEQ ID NO:5).
10. A protein of Claim 8 encoded by the amino acid sequence (SEQ ID NO:6).

-18-

11. A protein of Claim 8 encoded by the nucleotide sequence (SEQ ID NO:17).
12. A protein of Claim 8 encoded by the amino acid sequence (SEQ ID NO: 18).
- 5 13. A protein of human megakaryocytic origin which exhibits significant sequence homology with the FGF receptor family of protein tyrosine kinases.
14. A protein of Claim 10 encoded by the nucleotide sequence (SEQ ID NO:7).
- 10 15. A protein of Claim 10 encoded by the amino acid sequence (SEQ ID NO:8).
16. A protein of human megakaryocytic origin which exhibits significant sequence homology with the NGF receptor family of protein tyrosine kinases.
- 15 17. A protein of Claim 14 encoded by the nucleotide sequence (SEQ ID NO:11).
18. A protein of Claim 14 encoded by the amino acid sequence (SEQ ID NO:12).

-19-

19. A DNA expression vector containing a DNA sequence of human megakaryocytic origin which hybridizes to a DNA fragment which hybridizes to DNA encoding an amino acid sequence present in the catalytic domain of a protein tyrosine kinase of the c-kit subgroup of protein kinases.
20. The DNA expression vector of Claim 17 containing a DNA sequence selected from the group consisting of:
- a) SAL-S1 (SEQ ID NO:5);
 - b) SAL-D4 (SEQ ID NO:7);
 - c) LpTK 2 (SEQ ID NO:9);
 - d) LpTK 3 (SEQ ID NO:11);
 - e) LpTK 4 (SEQ ID NO:13); and
 - f) LpTK 13 (SEQ ID NO:15).
 - g) LpTK 25 (SEQ ID NO: 22 and
 - h) HpTK 5 (SEQ ID NO: 23).
21. A cell transformed by the expression vector of Claim 17.

1/35

sals1 (160 bases)
FLK1-LIKE

1	PTK1/3 PRIMERS	21	41
5' <u>ggatcctgtgcatcagtgacttagggctaggaacattctgtctcggaagcgacgtggt</u>			
		D P Y H Q A L R A R N I L L S E S D V V	
61		81	101
gaagatctgtgactttggccttggccgggacatctacaagagacccagctacgtccgcaa			
		K I C D F G L A R D I Y K D P S (Y) V R K	
121		141	
gcatacccggtgccctgaagtggatggcgccagaattc 3'			
		H A R L P L K W H A P E F	

FIGURE 1

sald4 (147 bases)-
FGFR-LIKE

1	PTK1/3 PRIMERS 21	41
5' <u>ggatccattcacagagac</u> cttagcagcacgcaacatcctggtctcagaggacctggtacc		
		G S I H R D L A A R N I L V S E D L V T
61	81	101
aagtcagcgactttggcctggccaaagccgagcggaagggttagactcaagccggtg		
		K V S D F <u>G L A K A E R K G L D S S R L</u>
121	PTKKW PRIMER 141	
<u>ccgtcaaatggatggctcccgaattc</u> 3'		
		P V K W H A P E F

FIGURE 2

3/35

LpTK2

GTTGGAATTCCTTCCGGCGCCATCCATTTACCGGCAGCTTTATTTTCGTGTCTAGATTCA
TAGATGTCTTCATTATCTACCTTAAAACTCTGGCAAGTCCAAAATCTGCTACTTTGTAG
ATATTATGTTCAACCAACGAGGACATTCCT

FIGURE 3A**LpTK3**

GTGCACAGGGATCTCGCGGCTCGGAACATCCTCGTCGGGGAAAACACCCCTCTCGAAAGTT
GGGGACTTCGGGTTAGCCAGGCTTATCAAGGAGGACGTCTACCTCTCCCATGACCACAAT
ATCCCCTACAAATGGATGGCCCCCTGAGGGAA

FIGURE 3B**LpTK4**

GTTACCGGAGATCTCAAGTCCAACAACATTTTGCTGCTGCAGCCCATTGAGAGTGACGAC
ATGGAGCACAAGACCCTGAAGATCACCGACTTTGGCCTGGCCCGAGAGTGGCACAAAACC
ACACAAATGAGTGCCGC

FIGURE 3C**LpTK13**

GTCAATCGTGACCTCGCCGCCCCGAAATGTGTTGCTAGTTACCCAACATTACGCCAAGATC
AGTGATTTCCGACTTTCCAAAGCACTGCGTGCTGATGAAACTACTACAAGGCCAGACC
CATGGAAAAGTGGCCTGTCAAGTGGTACGCTCCGGAATGCATCAACTACTACAAGTTCTCC
AGCAAAAGCGATGTCTGGTCCTTTGGAATTC

FIGURE 3D


```

1  TTCGAGCTCG CCGGACATTG ATTATGACT ATTATTAAT AGTAATCAAT TACGGGGTCA TTAGTTCATA GCCATATAT GCCATATAT GGAGTTCGGC GTTACATAAC
   AUGCTCGACC GGGCTGTAC TATTAAGTGA TCAATAATTA TCATTAGTTA ATGCCCCAGT ATCAAGATAT CGGGTATATA CCGTCAAGCG CAAATGTATTG

101 TTAGCGTAA TGUCCCGCCT GCGTACCGC CCAACGACC CCGCCCATTT AGTCATATA TGACGTATGT TCCCATAGTA ACCCAATAG GCATTTTCCA
   AATGCCATT ACCGGCGGA CCGACTGGC GGTGTCTGG GCGGGTAACT TGCAGTAT ACTGCATACA AGGTATCAT TCGGTATAT CCGTAANGGT

201 TTGACGTCAA TGGGTGGAGT ATTAGGTA AACTGCCAC TTGGCAGTAC ATCAAGCTA TCATATGCCA AGTACGCC CATTACAGT CAATGACGGT
   AACTGCAGT ACCGACTCA TAAATGCCAT TTACCGGTG TACCTCATG TAGTTCACAT AGTATACCGT TCATGGCGGG GATAACTGCA GTTACTGCCA

301 AATGGGCCG CCTGGCATTA TGCCGAGTAC ATGACCTTAT GGGACTTTCC TACTTGGCAG TACATCTAGC TATTAGTCAAT CGTATATACC ATGTGATGCC
   TTTACCGGC GACCCGTAT ACGGTGATG TACTGCAATA CCGTGAAGG ATGAACCGTC ATGTAGATGC ATATATCAGTA CGGATAATGG TACCAGTACG

401 GGTTTTGGCA GTACATCAAT GGGGTGGAT AGCGGTTTGA CTCACCGGA TTTCCTAATC TCCACCCCAT TGACGTCAAT GGGAGTTGT TTTGGCACCA
   CCAAAACCGT CATGTAGTGA CCGGACCTA TCGCCAACT GAGTGGCCT AAGGTTCAG AGGTGGGTA ACTGCACTTA CCGTCAATCA AAACCGTGGT

501 AATCAACGG GACTTTCCAA ATGTGCTAA CAATCCGCC CCAATGAGC AATGGCGG TAGCGTCTA CGGTGGGAGG TCTATATAG CAGAGCTCGT
   TTTAGTTGCC CTGAAGGT TTACAGCAT TTGAGGCGG GTTACTGGG TTTACCGGC ATCGGCACAT GCCACCTCC AGATATATTC GTCCTCGACCA

601 TTAGTGAACC GTGAGATCC CTGGAGACG CATCCACCT GTTTGACCT CCAATAGAGA CACCGGACC GATCCAGCTT CCGCGCGCGG GAACGTGCA
   ATCACTTGG CAGCTAGCG GACCTCTGG GTAGGTGGA CAAACTGGA GGTATCTCT GTGGCCCTGG CTAGGTGGA GCGCGCGGCC CTTCGCAGCT

```

FIGURE 4A

5/35

701 TTGGAACGGG GATTCCCGGT GCCAAGAGTG ACGTAAGTAC CGCCTATAGA GCTATAGGC CCALTTGGCT TCCTTAGAAC CGGCTACAA TTAATACATA
AACCTTGCGC CTAGGGGCA CGGTTCTCAC TGCATTCTATG GCGGATATCT CAGATATCCG GGTCAACCGA AGCAATCTTG CGCCGATGTT AATTATGTAT

801 ACCTTATGTA TCATACACAT ACGATTTAGG TGACACTATA GANTACATC CACTTTGCTT TTCTCTCCAC AGGTGTCAC TCCCAGGACC AACTGCACCT
TGGAAATCAT ACTATGTGA TGCTAATCC ACTGTGATAT CTTATTCTAG GTGAACGGA AAGAGAGGTG TCCACAGGTG AGGTCCAGG TTGACGTGGA

901 CGGTTCTATC GATTGAATC CCGGGGATC CTCAGAGAT CCGTGACCT CGAGATCCAT TGTGCTGGCG CGGATTCCTT ATCACTGATA AGTTGGTGGA
GCCAAGATAG CTAACCTAAG GGGCCCCCTAG GAGATCTTA GGAAGCTGGA GCTCTAGGTA ACACGACCGC GCTTAGAAA TACTGACTAT TCANCCACCT

1001 CATATTATGT TATACAGTGA TAAAGTGTA ACCATGACAA AGTGCAGGC GAATACAGTG ATCCGTGCGG CCGTAGACCT GTTGACGAG GTCCGGCGTAG
CTATATATCA AATAGTCACT ATTTCACAGT TCCTACTGTT TCAACGTCGG CTTATCTCAC TAGGCACGGC GGGATCTGGA CAACTTGCTC CAGCCGCTC

1101 ACGGTCTGAC GACAGGCNA CTGGCGGAAC GGTGCGGGT TCAGCAGCGG GCGTTTACT GGCACCTCAG GAACAGCGG GCGCTGCTCG ACGCACTGCG
TGCCAGACTG CTGTGCGTT GACCGCCTTG CCAACCCCA AGTCGTCGGC CGCGAATGA CCGTGAAGTC CTTGTTGCGC CGCAGCAGC TCGGTGACCG

1201 CGAACCCTATG CTGGGGAGA ATCATAGCAC TTGGTGCGG AGGCCGACG ACGACTGGCG CTCATTCTTG ACTGGGANTG CCGCCAGCTT CAGGCAGGGG
GCTTCGGTAC GACCGCCTCT TAGTATCGTG AAGCCACGGC TCTCGGCTGC TCGTGACCGC GAGTAAGAC TGACCCCTTAC GGGCGTCGA GTCCGTCCCG

1301 CTGCTCGCT ACCGCCAGCA CAATGGATCT CGAGGATCT TCCATACCTA CCAGTTCTGC GCGTGCAGGT CCGGGCCCA CTACTCTTG ATGTATTACT
GACGAGCGGA TGGCGGTGCT GTTACCTAGA GCTCCCTAGA AGGTATGGT GGTCAAGACG CGACGTCCA GCGCCGGCT GATGAGAAAC TACATANTGA

FIGURE 4B

1401 CATATTACCA AGGAATAACT GGCGGGCACA GGTCAGGTG CTGAAGGGAC ATTGTGAGAA GTGACCTAGA AGGCAAGAGG TGAGCCCTCT GTACACGCTG
 GTATATGGT TCCTTATTGA CCGCGGTGT CCGAGTCCAC GACTTCCCTG TAAACACTCT CACTGGATCT TCCGTCTCC ACTCGGAGGA ACTCGCGACC

1501 CATAGGGGC GGTGAGGGC TCTTGGTCA AGCAGTAACG CCACTGTCTG GGAAGGCACC TGTTACTCAG CAGACCATGA AAGGGCTCT CCCTTTCCTT
 GTATCCCG CGAATCCCG AGAATCCCG TCGTCAATCC GGTCAACAGC CTTCCCTGG ACAATGATC GTCTGTACT TCCCGCAGA GGGAAAGGAA

1601 GCAGCACTCA GGGAACTCT TCTCCACCA GCTTCTGTG GGAGGTCTGA TATTATCAG GCTGCCCCG AGTCATCCG AGCCTAACG CCTCCCTCTG
 CCTCCTCAGT CCTTGTGAG ACGAGTGGT GGAAGAACAC CCTCCGACCT ATANTAGTC CGGACGGCG TCAATAGGCC TCCGATTCG GCAGGGACAC

1701 GIGTTTCAGT GGTACACTC CTGTCCACT TCTATGCTC TCTTGGCTC CTTGTCTCTG TCTTCTCTG CTTCAGCTAC CTGAAGCCG TTTCTGTCT
 CACGAACTCA CCACTGTGAG GAACAGTGA AATACAGG AGAACCGAG ACCAGCGGAG AGAGAGAGAC GAATCGATG GACTTCGGG AAGAACAGA

1801 CTTAAGTCT TTGATCTTC TTATAAGTGC AGAGAAGAA TCTGACGTA TCTGCTCTC TCTTCTCTG CTTCAGCTAC CTGAAGCCG TTTCTGTCT
 GAATTCAGA AACTAGAAAG AATATTCAGG TCTCTCTCTT ACCACTGCAT ACCAGCGGAG AGAGAGAGAC GAATCGATG GACTTCGGG AAGAACAGA

349

1901 ATACCTGCTC TCTATCTGT CACACTCTC CGAGGCCAGC ACCATCCAC TGTCTGTCTG GTTGTCCACA GAGCCTTCT AGTCTGTAGG GGTCTATGGG
 TATGACGAG AGATAGACGA GTCTGAGGAG GCTCCGGTCT TGTAGGTG ACAGACAGC CAACAGGTCT CTCGGAACA TCCAGCAAC CCAGTACCC

341 Y R S E I Q E C E E S A L V M C S D T Q N D V S Q K Z T T P Z H P

2001 ATTCTCTCA ATGTCTTCT CTTGAGGAA CCAAGGTCT CAGCCCTCT GGCAGGGAC CCGGAAGG ACACCAAGT GTATPACTG GCGCCAGG
 TTAGGAGCTT TACAGAGTA GGACTCTCT GTTCCCGA GTCCGGAGA CCGTCCCTG GGCCTTCC TGTGGTCA CATATGGAC CGCGGTCTG

308 P E E F T K M R S S G R T E A C R A L C G P F S V W N Y Y R A A L

FIGURE 4C

2101 TGTGGCGCTG CAGGCTTGGC GGGTGTCTT CAGCGTCAGC CTGGGGCCA TGTGGACAC CTGGGAGAG CTGCCCTCTT CTGAGCTCTG
 ACACCGGAC GTCCGACCG CCGGACAGG GTCCGAGTGC GACCGGTAC ACATCCGGT ACCACTGTG GACGCTCTC GACGGAGAA GACTCGAC
 275 S H R Q L S P P S D E A D A Q A I H L A M T S V Q S F S G E S S Q

2201 AGAGCTGGC GGGGCCATGC AGACTCTCTC TTCTCTTGC AGGCCCTGC CTGGAGCAG GTCCCCCAGG ATCTCCACCA GCTCCGAGAA TGCAGCTCTC
 TCTCGACCG CCGCGTACG TCTGGAGAG AAGGAGACG TCCGGGACG GGACTCTG GAGGGTCC TAGAGTGT CGAGGCTCTT ACCTCCAG
 241 S S R P A M C V E E E Q L G R G O L L D G L I E V L E S F A P R

2301 GCCTTGGGT CTCCGGACCA GCACTACAGC ATGATGGGC GATGGCGG AGTGGCCAGC TCCGGGCCC TCATCTTCTT GCCGCTCTC AGCCGCTGGC
 CGGAACCCA GAGGCTGT CGTCAAGTCG TACTAGCCG CATACGCC CATACGGTCG TACCGGTGC AGGCCCGG AGTAGGAACA CGGCAGAGAG TCGGCGACCG
 208 A K P D G S W C N L M I R R I A P T A L E P A R M R T G D R L R Q

2401 AGACTCTCTC ATGATCTGC ACCCAGGCT ACGGGAGCC CCGCAGAG AGATCTCC AGAGAGCAG CCGAAGAC CACAGTCA TCTGGTGT
 TCTGAGGAG TAAGTACG TGGGTCCA TGCCCTCG GGGTCTCT TCTGAGG TCTCTCTG GGTCTCTG GTCTCTCTG GTCTCTCTG AGACGACCA
 175 C F E E N I Q V G P Y P S A G L S F I E W L L V G F S W V D S Q T T

2501 GTACACUTTG TCGAAGATGC TTTCAGGCG CATCCACTTC AGGGCAGCC GGGCACTGCC CTGGCGAGC TAGTCGGGT CTTCGAGAT GTCCCGGCA
 CATGTGAC AGCTTCTACG AAGTCCCG GTAGGTGAG TCCCGTCCG CCGGTGACG GAACGCTGC ATCAGCCCCA D K Y I D R A
 141 Y V K D F I S E P A M W K L P L M A S G K R V Y D P

2601 AGCCCAAGT CACAGATCT CACACAGTGC CTTTCGACA GCAGATGTT CCGAGCAGC AGGCTCTGT GGTGACTT TCGGAGCC AGGAATCCA
 TCGGTTTCA GTGTCTAGA GTGGTGCAG GARAGGCTGT CGTCTACA GGTCTCTCG TCCAGAGCA CCTACGTGA AGCCCTCTGG TCTTGAGGT
 108 L G F D C I K V V D S E S L L I N R A A L D R H I C K R S A L F E

2701 TCCCTCTGCG CACCTGGAG CTGTAGCAGA CAAGATCTC CATGTCTAGC GGGCTCAGC ACAGGTCTCT AGCTCTCTG TCTGGAGAG CCGGCTCTG
 AGGAGACCG GTGGACCTT GACATCGTCT GTCTAGAG GTACAGTGC CCGAGTCCG TGTCCAGAG TCGAAGAC AGACTCTCTT GGGCGAGCG
 75 M G R A V Q F S Y C V L D E M T L P S L W L D E A E Q D P S A R R A

FIGURE 4D

FIGURE 4E

2801 TCCGCCCTCG GTCTTCUNGA ACCGUGGAA GAGGACCTG TCGTGTCTC CCGCCGCCCT CCGATCCAGC CTGGCGAGCT CCACCATGGC GCGGAAGCGT
 AGCGGGGAGC CAGAGCTCT TGGGGCCTT CTCTGGGAC AGCAGCAGG GCGCGCGGA GGTAGGTGG GACCGTCA GGTGTACCG GCGCTTCCCA
 41 C G E T K S P R A P L V R D S S G P R R R A L E V M A R P R
 2901 CCGCGGTGCT CCGGAGACTT CTCTGGCGA TGCACGAGC TGGCTGAGG GCGCCAGTC GTCCGCCGA GAGCGGCTC CATTCGCCCG CCGCCCGCG
 GCGCGGACGA GCGCTCTGA GAGGAGCCT AGGTGCTTG ACCGAGCTC CCGCGTCAG CAGCGCGGT CTCCCGGAG GTAAGGGGCG GCGCGCGCGC
 8 G R Q E P S K E
 3001 CCGCGGCGC GCGGCCGCT CCGCGGCGC GCGCGGCGC GCGCGGCGC GCGCGGCGC GCGCGGCGC GCGCGGCGC GCGCGGCGC GCGCGGCGC
 GCGCGGCGC GCGCGGCGC GCGCGGCGC GCGCGGCGC GCGCGGCGC GCGCGGCGC GCGCGGCGC GCGCGGCGC GCGCGGCGC GCGCGGCGC
 3101 CTTATATAGG TTACAAATAA AGCAATAGCA TCACAAATTT CACAAATAA GCATTTTTT CACTGCATTC TAGTGTGCT TTGTCCAAAG TCATCAATGT
 GAATATACC AATGTTTATT TCGTTATCGT AGTGTATAA GTGTATATZ CGTAAATAA GTGACGTAG ATCAACACCA ACAGGTTTG AGTAGTACA
 3201 ATCTTATCAT GTCTGGATCG ATCGGGATTT AATCGGGCG AGCACCATGG CCGCAATAA CCTCTGAAG AGCAACTTGG TTAGGTACCT TCTGAGGCGG
 TAGAATAGTA CAGACCTAGC TAGCCCTTAA TTAAGCGCG TCGTGGTACC GGACTTTATT GGACACTTTC TCCTTGAACC AATCCATGGA AGACTCCGCC
 3301 AAAGAACCA GCTGTGGAATG TGTGTAGTT AGGTGTGGA AATGCCGAG GCTCCGAGC AGCAGAGAT ATGCAAGCA TGCATCTCAA TTAGTCAGCA
 TTCTTGTGC CACACCTTAC ACACAGTCA TCCACACCT TCCAGGGGTC CGAGGGGTC TCCGTCTCA TACGTTTCT ACCTAGGTT AATCAGTCT
 3401 ACCAGGTGT GAAATCCCC AGGCTCCCC GAGGCGAGAA GTATGCAAG GTATGCTC CATGCTATC AATAGTCAG CAACCATAGT CCGCCCGCTA ACTCCGCGCA
 TGTCCACAC CTTTACAGGG TCGAGGGGT CGTCCGCTT CATACGTTT GTACGTAGG TTAATCAGT GTTGATCA CCGCGGGGAT TCAGGCGGGT

9/35

FIGURE 4F

3501 TCCGCGCCCT ANCTCCGCC AGTTCGGCC ATTCTCCGCC CCAATGCTGA CTATTTT TTTATATGC AGAGGCGAG GCGCCTCGG CCTGTAGCT
 AGGCGGGGA TTGAGGGGG TCAAGGGGG TAAAGGGGG GTACCGACT GATTAANAA ANTAANATACG TCTCGGCTC GCGCGAGCC GGAGACTCGA

3601 ATTCCAGAG TAGTGAGGAG GCTTTTGG AGGCTAGGC TTTTGCNAA AGCTGTAAAC AGCTTGGCAC TGGCGTGGT TTTACAACT TTTGACTGGG
 TAAGTCTTC ATCAGTCTC CGAANANACC TCCGATCCG AAACGTTTT TCGACAAATG TCGNACCGTG ACCGACCA AATGTGGCA GCACTGACCC

3701 AAAAACTGG GGTATCCCA CTTAATGCC TTGACGACA TCCCCCTTC GCGAGCTGG GTAATAGCGA AGAGGCGCG ACCGATCGC CTTCCTCAACA
 TTTGGGACC GCANTGGGT GANTTAGCG AACGTCTGT AGGGGGGAG CGTCCGACC CATATCGCT TCTCGGGCG TGGTAGCGG GAAGGGTTGT

3801 GTTGGTAGC CTGAATGGG ANTGGGCT GATGGGTAT TTCTCTTA CGCATCTGG CGTATTTCA CACGCTAC GTCAAGCGA CCATAGTAGC
 CACGCATCG GACTTACCG GACTTACCG TTACCGGA CTACGCATA AAGAGGAA GCGTAGCAC GCGTAAAGT GTGGGTATG CAGTTCTGT GGTATCATGC

3901 CGCCCTGTAG CGCGCATTA AGCGGGCG GTGTGGTGT TACGGGAGC GTGACGCTA CACTGGCTA CACTGGCTA CACTGGCTA CACTGGCTA CACTGGCTA
 GCGGGACATC GCGCGTAA TCGCGCGCC CACACCGCC CACACCGCC CACACCGCC CACTGGCTA CACTGGCTA CACTGGCTA CACTGGCTA CACTGGCTA

4001 CCTTCTTT CTGCGCACGT TCGCGGCT TCGCGGCT TCGCGGCT TCGCGGCT TCGCGGCT TCGCGGCT TCGCGGCT TCGCGGCT TCGCGGCT
 CGAAGGAA GAGCGGTGA AGCGGCGA AGCGGCGA AGCGGCGA AGCGGCGA AGCGGCGA AGCGGCGA AGCGGCGA AGCGGCGA AGCGGCGA

4101 AAAAACTTG ATTGGCTGA TGGTTCACGT AGTGGGCGA GCGCGTATA GCGCGTATA GCGCGTATA GCGCGTATA GCGCGTATA GCGCGTATA
 TTTTGTGAC TAACCCACT ACCAAGTGA TCACCGGTA GCGCGTATA GCGCGTATA GCGCGTATA GCGCGTATA GCGCGTATA GCGCGTATA

FIGURE 4G

4201 TCTTCTTCCA AACTGGAACA ACACTCAACC CTATCTCGGG CTATCTTTT GATTATTAAG GGATTTGGC GATTTCGGC TATTTGGTAA AAATGAGCT
 AGAACAAAGT TTGACCTTGT TGTGAGTTGG GATAGAGCCC GATANGAANA CTAATATATC CTAATAACGG CTAAAGCCG ATAACCAATT TTTACTCGA

4301 GATTTAACAA AAATTAAACG CGAATTTTAA CAAATATTA ACCTTTACAA TTTATGGTG CACTCTCAGT ACAAATCTGCT CTGATGCCCG ATACTTAAGC
 CTAATTTGTT TTAAATATGC GCTAAATATT GTTTATANT TGCANATGTT AATAATCCAC GTGUAAGTCA TGTTAGACGA GACTACGGCG TATCAATTCG

4401 CAACCTCCGT ATCGCTACGT GACTGGGTCA TGGCTGGCC CGACACCCG CCAACACCCG CTGACGGGCG CTGACGGGCT TGTCTGCTCC CGGCATCCG
 GTTGAGGCGA TAGCGATCCA CTGACCCAGT ACCGACGGCG GGCTGTGGC GGTGTGGCG GACTGCCCG ACAGACGAGG GCCGTAGCG

4501 TTACAGACAA GCTGTGACCG TCTCCGGGAG CTCATATGCT CAGAGGTTTT CACCGTCAATC ACCGAAACG CGAGGCGAGT ATCTTGAG ACCAAAGCC
 AATGTCGTT CGACACTGGC AGAGGCCCTC GAGTACACA GTCTCCAAA GTGGCAGTAG TGGCTTGG CGCTCGTCA TAAGAACTTC TGTCTTCCC

4601 CTCGTGATAG GCTATTTTT ATAGGTTAAT GTCATGATAA TAATGCTTTC TTAGACGTCA GTTGGCAGTT TTGGGGGAAA TGTGGCGGA ACCCTATTT
 GAGCACTATG CGGNTAATAA TATCCAAATA CAGTACTATT ATTACCAAG AATCTGCAGT CCACCGTGA AGCCCCCTT ACACGGGCT TGGGGATAA

4701 GTTATTTTT CTAATACAT TCAATATGT ATCCGCTCAT GAGCAATAA CCTGTATAA TGTTCATA TATTTGAAA AGGAAGATTA TGATATTCA
 CAAATAAANA GATTATGTA AGTTATACA TAGGCGAGTA CTCTGTTAT GGGACTATTT ACGAAGTTAT TATAACTTTT TCTTCTCAT ACTCATAGT

4801 ACATTTCCGT GTCGCCCTTA TTCCCTTTTT TGGGCGATTT TGGCTTCTG TTTTGTCTA CCCAGAAACG CTGGTGAAG TAAAGATCC TGAAGATCAG
 TGTAAAGGCA CAGCGGGAAAT AAGGGAAAA ACGCCATAA ACGBAAGGAC AAAACGAGT GGTCTTTGC GACCACCTTC ATTTCTACG ACTTCTAGT

FIGURE 4H

4901 TTGGGAGCAG GAGTGGGTTA CATCGAAGTC GATCTCAACA GCGGTAAAGT CCTGAGAGAT TTTCGCCCGG AAGAACGTTT TCCATGATG AGCACTTTT
 AACCCACGTG CTCACCCCAAT GTAGCTTGAC CTAGAGTTGT CGCCATTCTA GGAATCTCTA AAGCGGGGC TTCTTGCAAA AGGTACTAC TCGTGAANA

5001 AAGTCTGCT ATGTGGCGG GTATTATCC GTGATGACCG CGGCAAGAG CAACTCGGTC GCCGCATACA CTATTCTCAG ATGACTTGG TTGACTACTC
 TTCAAGACGA TACACCGCGC CATATATGG CACTACTGG CCCCCTCTC GTTGAGCCAG CGCGTATGT GATNAGATC TTACGAAACC AACTCATGAG

5101 ACCACTCACA GAAAGCATC TTACGGATGG CATGACACTA AGAGATTTAT GCAGTGTGG CATTAACCATG AGTGAZACA CTGGGCCAA CTTACTTCTG
 TGGTCACTGT CTTTCTGTG AATGCCATC GTACTGTCTT TCTCTAATA CGTCACGACG GTATTGGTAC TCACTATTGT GACGCGGTT GAATGAAGAC

5201 ACAACGATCG GAGGACCGAA GGAGCTAACC GCTTTTTCG ACAACATGG GATCATGTA ACTGCGCTTG ATCGTTGGGA ACGGAGCTG AATGAAGCCA
 TCTTGCTAGC CTCCTGGCTT CCTCGATTGG CGAANAANAAC GCTTTTTCG TGTGTATACC CCTAGTACAT TGAGCGGAAC TNGCAACCT TGGCTCTGAC TTACTTCTGT

5301 TACCAACGA CGAGCGTGAC ACCACGATGC CAGCAGCAAT GGCAACAAG TTGGCAAC TATTACTGG CGAATCACTT ACTTAGCTT CCGGCCAACA
 ATGGTTTGT CCTCGCACTG TGGTGTACG GTCTGTCTTA CCGTTGTTC AACCGTTTG ATATTGACC CTTGATGNA TGAGATGNA TGGCGCTGT

5401 ATTAATAGAC TGGATGGAG CGGATTAAGT TGCAGGACCA CTTCGCGGT CGGCCCTTC GGTGGCTGG TTATTCTCTG ATAACTCG AGCGGTGAG
 TAATTATCTG ACCTACCTCC CCTATTCTA ACCTCTCTGT GAAGACCGA GCGGGGAGG CCGACCGACC AATNACGAC TATTAGAC TCGGCCACTC

5501 CGTGGGCTC GCGGTATCAT TGCAGCACTG GGGCCAGATG GTAGCCCTC CCGTATCGTA GTTATCTACA CCACGGGGAG TCAGGCNACT ATGATGNAAC
 GACCCACAG CGCCATAGTA AGTCTGTGAC CCCGTCTAC CATTCGGAG GGCATAGCAT CAATAGATGT GCTGCCCTC ACTCGTTGA TACCTACTTG

FIGURE 4I

5601 GAAATAGACA GATCGCTGAG ATAGGTGCT CACTGATTAA GCATTGGTAA CTGTGAGACC AAGTTTACTC ATATATCTT TAGATTGATT TAATACTT
CTTTATCTGT CTACCGACTC TATCCACGGA GTGACTAATT CGTAACCATY GACAGTCTGG TTCAATGAG TATATATGAA ATCTAATCTAA ATTTGAAK

5701 TTTTITATTT AARAGATCT AGTGAAGAT CCTTTTIGAT AATCTATGA CCAAAATCC TTAAGTGAG TTTTCTTCC ACTGAGCTC AGACCCCGTA
AAAAATTAAA TTTTCTAGA TCCACTICTA GGAATACTA TAGAGTACT GTTTTAGG AATTCCACTC AANAAGCAAG TCACTGCGAG TCTGGGGCAT

5801 GAAAAAATCA AUGATCTTC TTGAGATCTT TTTTCTTGC GCGTAATCTG CTGCTTGCA ACAAATAAAC CACCGCTACC AGCGTGCTT TGTTCGCGG
CTTTCTAGT TTCTTAGAAG AACTCTAGGA AANAAGACG CGCATTAGAC GACGAACGTT TCTTTTITG GTGCGATGG TCGCCACCAA ACAATCGGCC

5901 ATCAAGAGCT ACCACTCTT TTTCGAGG TAACTGGCT CAGCAGAGCG CAGATACCAA ATACTGTCTT TCTAGTGTAG CCGTAGTTAG GCCACCACTT
TAGTCTCGA TGGTTGAGAA AAGGCTTCC ATTGACCGAA GTCGTCTGCC GTCATGGTT TATGACAGGA AGTACATC GGCATCAATC CCGTGTGAA

6001 CAAGAACTCT GTAGCACCGC CTACATACCT CGCTCTGCTA ATCTGTATC CAGTGGCTGC TGCCACTGCC GATAGTCTT GTCTTACCGG GTTGGACTCA
GTCTTTGAGA CATCGTGGC GATGTATGGA GCGAGACGAT TAGGACAATG GTACCGACG ACGGTACCG CTATTCAGCA CAGATGGCC CAACCTGAGT

6101 AGACGATAGT TACCGGATTA GCGCCAGCG TCGGGCTGAA CCGGGGCTC GTGCACACAG CCCAGCTTG AGCGAAGAC CTACACCGAA CTGAGATACC
TGTGCTATCA ATGGCTATT CCGCGTGGC AGCCCGACTT GCCCCCAAG CACGTGTCT CCGTGTCTG GATGTGGCTT GACTCTATGG

6201 TACAUCGTGA GCATTGAGAA AGCGCCAGC TTCCCGAAG GAGNAAGCG GACAGGTATC CGGTAAAGCG CAGGCTCGA ACAGGAGAGC GCACGAGCG
ATGTCCGACT CGTAACCTT TCGCGGTGG AAGGCTTCC CTCTTTCGC CTGTCCATAG GCCATTCGCC GTCCAGCTT TGTCTCTCG CGTCTCTCC

FIGURE 4J

6301 GCTTCCAGGG GGAACCCCT GGTATCTTA TAGTCTGTC GGGTTGCGC ACCTCTGACT TGAGGTCGA TTTTGTGAT GCTGCTCAGG GCGGCGGAGC
CGAAGGTCCC CTTTGGCGA CCATAGAAAT ATCAGGACAG CCAAGCGG TGGAGACTGA ACTGCGAGCT AAAACACTA CGAGCAGTCC CCGCGCCTG

6401 CTATGGAAA ACGCAGCAA CCGGCGCTTT TTACGGTTC TGGCTTTTG CTGGCCTTTT GCTCAGATG TCTTCTCTCC GTTATCCCT GATTCTGTGG
GATACCTTTT TCGGTCTGTT GCGCGGAAA ANTGCCAAG ACCGGAAC GACCGMAA CGAGTGATCA AGAAGGAGC CAATAGGGA CTAGACACC

6501 ATACCGTAT TACCGCCCTT GAGTGAGCTG ATACCGCTG CCGAGCCGA ACCACCGAGC GCAGCGATC AGTACGGAG GAACCGGAG AGCCCCAAT
TATTGGCATA ATGGCGAAA CTCACTCGAC TATGGCAGC GCGCTGGCT TGTGGCTCAG TCACTGCTC CTTCGCTTC TCGCGGTTA

6601 ACBCAACCG CCTCTCCCG CGCGTTGGC GCUCNACCG CTAACTAAT AGTTCGACG TCCAGCTGGC ACGACAGTT TCCGACTGG AAGCGGCA GTGAGCGCA CGCAATTAT
TCCGTTTGGC GACAGGGGC GCUCNACCG CTAACTAAT AGTTCGACG TCCAGCTGGC ACGACAGTT TCCGACTGG AAGCGGCA GTGAGCGCA CGCAATTAT

6701 GTGAGTACC TCACTCATA GGCACCCAG GCTTACACT TTATGCTTC GGTCTGATG TTGTGGAA TTGTGAGCG ATACATTT CACACAGGA
CACTCANTGG AGTGAATAT CCGTGGGTC CGAATGTGA ATACGAGG CCGAGCATAC AACACCTT AACACTGCC TATGTANA GTGTGCTT

6801 ACAGCTATGA CCATGATTAC GAATTAA
TGTCGACT GGTACTAATG CTTAAT

14/35

FIGURE 5A

1 TTCCAGACTCG CCGGACATATG ATATATATGACT AGTATATATATAT TACGGGCTCA TTAGTTTCATA GCCCATATATAT GGAGTTCCGC GTTACATATAC
AAGCTCGAGC GGGCTGTATAC TAATATATATG TAATATATATG TCAATATATATG TCAATATATATG TCAATATATATG TCAATATATATG TCAATATATATG

101 TTACGGTAA TGGCCCGCCT GGCTGACCGC CCACGACCC CCGCCCATATG ACCTCAATAA TGACGTATGT TCCCATATATG AGCCCAATATG GGACTTTCCA
AATGCCATTT ACCGGGCGGA CCGACTGGCG GGTTCCTGGG GCGGGTATAC TGCAGTTATT ACTGCATACA AGGTATCAT TGGGTATATC CCTGAAAGGT

201 TTGACGTCMA TGGGTGAGT ATTTACGGTA AACTGCCAC TTGGCAGTAC ATCAAGTATA TCATATGCCA AGTACGCCCC CTATTGACGT CAATGACGGT
AACTGCAGTT ACCCAGCTCA TAAATGCCAT TTGACGGGTG AACCCTCATG TAGTTTACAT AGTATACGGT TCATGCGGGG GATAACTGCA GTTACTGCCA

301 AAATGGGCG CCTGGCATTA TGCCAGTAC ATGACCTTAT GGGACTTTCC TACTTGGCAG TACATCTACG TATTAGTCTAT CGCTATTACC ATGTTGATGC
TTTACCGGCG GGACCGTAT ACGGTCTATG TACTGGATAT CCTGAAAGG ATGAAACCTC ATGTAGATGC ATAATCAGTA GCGATAATGG TACCACCTACG

401 GGTTTTGGCA GTACATCAAT GGGGTGGAT AGCGTTTGA CTCACGGGA TTTCCAAGTC TCCACCCCAT TGACGTCAAT GGGAGTTTGT TTGGCACCA
CCAAACCCGT CATGTAGTTA CCGGCACCTA CCGCCAACT TCGCCAACT GAGTCCCCCT AAAGTTTCAG AGGTGGGATA ACTGCAGTTA CCTCAAAACA AAACCGTGGT

501 AAATCAACGG GACTTTCCAA AATGTGTTAA CAATCCGCC CCATTGACGC AAATGGGCG TAGGCTGTGA CGGTGGGAGG TCTATATATG CAGAGCTCGT
TTTAGTTGCC CTGAAGGTT TTACAGCATT GTTGAGGCG GTTAACTGCG TTTACCCGCC ATCCGCACAT GCCACCCCTCC AGATATATTC GTCTCGAGCA

601 TTAGTGAACC GTGAGATCGC CTGAGAGCG CATCCACGCT GTTTTGACCT CCATAGAGA CACCGGAGC GATCCAGCCT CCGCGGCCCG GAACGGTGCA
AATCAGTTGG CAGTCTAGCG GACCTCTGCG GTAGGTGCG CAAAACTGGA GGTATCTTCT GTGCCCCCTG CTAGGTGCGA GCGCGGCCGC CTTGCCACGT

701 TTGGAAACCG GATTCGCCGT GCCAAGAGTG AGTATAGTAC GGCCTATAGA GTCTATAGCG CCATTTGGCT TCGTTAGAAC CCGGCTACAA TTAATACATA
AACCTTGGCG CTAAAGGCGA CCGTTCTCAC TGCAATTCATG CCGGATATCT CAGATATCCG GGTGAACCGA AGCAATCTTG CGCCGATGTT AATTATGTAT

FIGURE 5B

801 ACCTTATGTA TCATACACAT ACGATTTAGG TGACACTATTA GATACATC CACTTTGCCT TTCTCTCCAC AGGTGTCCAC TCCAGGTCC AACTGCACCT
 TGGATATACAT AGTATGTGTA TGCTAAATCC ACTGTGATAT CTTATTGTAG GTGAAACGGA AAGAGAGGTG TCCACAGGTG AGGTCCAGG TTGACGTGGA

901 CGGTTCTATC GATTGAATTC CCGGGGATC CTCTAGAGAT CCCTCGACCT CGAGTCGACT TTTTGTGAG CCAAGGGTA CTCTTTTTC
 GCCAAGATAG CTNACTTAA GGGCCCTTAG GAGATCTCTA GGGAGCTGGA GCTCAGCTGA AAAAANAANA AAAACATCC GTTTTCCCAT GAAGNAANAAG

1001 TTTATTAAAT ACTCAGAAGT CTAGGCCACA GCAATCTACT GTTCTCTCT CTATTTCCTA AACTATTTTG ATACCTATTT CTCAGACTTT ATGGGCTATT
 AATAAATTA TGAGTCTTCA GATCCGGTGT CGTTAGATGA CAAGAGGAGA GTAAAGGAT TTGATNAAAC TATGGATAAA GAGTCTGAAA TACCCGATNA

1101 AGACATTCT CACATTTCOA TAGATAATAA CTCATCCGTT TTGCAACCTG ATTCCTAATA TTAAGAGATT AAACTAATG TATATGACTC TCAGTTGACA
 TCTGTAAAGA GTGTAAAGGT ATCTATTATT GAGTAGGCA ACGTTGGAC TAAGAGTTAT ATTCTCTA TTTTGATTAC ATATACTGAG AGTCMACTGT

1201 CATACTGAAG TACAGAAAAA TTCCATCAAT TCCTCTGCA AATGAAAAA GACTTCGTT TCTCAACAGC TGCATCATTT TTTATGCAAT AGAANAAT
 GTATGACTTC ATGTCTTTT AAGGTAGTAA AGGAGACGT TTTACTTTTT CTGAAGCANA AGAGTTGTG ACGTAGTAAA AAAATACGTA TCCTTTTTTA

1301 GTGCAATTAC TCCAAGTACA ATCAAGTCAT TTAACATGGC TTTACCATCA TTGTAGTTAC AGGATATTTT AAAAGAGAAA AAAAATCTC AAAGCACAGG
 CACGTTAATG AGGTTCAATG TAGTTCACTA AATTGTACCG AANTGGTAGT AACATCAATG TCCATATAAA TTTCTCTTT TTTTGTAGG TTTCGTGTCC

1401 TCCTGCTGTG CAGCAAGCA ATCAAAATCC TTCATAATTA CAGCCTGATG GGATTCAGCA ATCTGAGGAA TAATGAATAA CCACCTAAT CAGTAAACAG
 AGGACGACAC GTCGTTTCGT TAGTTTAAAGG AAGTATTATT GTGGGACTAC CCTAAGTCGT TAGACTCCTT ATTACTTATT GGTGAGATTA GTCATTGTG

1501 GAANATGCTA CAACAGTCAC TGAGTAAAAA TTGAGCTATC ATCTGTGAT TCCTTGATC GACATTTCAA ACNATAATG GAATGTAAAG TATCTCTTNA
 CTTTTACGAT GTGTGAGTG ACTCATTTTT AACCTGATG TAGACAATA TAGAAGTAC CTTTACATTC ATAGAGATTT

FIGURE 5C

1601 AAGAGAAAT AACTTGGTTT AGTGTGCTTA ATTTTACCAG GCAGTGAGGA AATTATATAT CACCTTGACT GTCTGCACT GTTGCCCACT CAATAAATG
 TTCTTTTTT TTGAACCAAA TCACACGAAT TAAATGGTC CGTCACTCCT TTAATATATA GTGGAACCTGA CAGGACCTCA CAACGGCTCA GTTATTTTAC

1701 CACAAATAT CTTTTTCATA ATACATGGCC AACTTTATCC TATCATTGTA ATATGTCAGG ATAACTGAT TGTGCAGTTG GTTGATAACA TTGTATTTTG
 GTGTTTATTA GAAAAAGTAT TATGTACCGG TTGAATATAGG ATAGTGAAT TATACAGTCC TATTTGACTA ACACGTCAAC CAATATTTGT AACATAAAC

1801 GAATGGATTA TTGGAATTTG TTTTGCTACT TTATTTATTTG ATATCTTCTT CCAGTGTTCA TCATTATGAG TTAATGTCAT CTGAATATGA AGATCTGTT
 CTTACCTAAT AAACCTTAAC AAACGATGA AAATATTAAC TATAAGNAGA GGTCACAAGT O R I F N N A D S Y S S D T
 506

1901 TCAAAATAGT CTTCAGTTT CCAACGCACT GTCTCAAAATG TAGGTGTTTCT CTTAGGCTCT GCATTCCAGC ACTCCAACAT GATGTTGTA AATGCTGTTG
 AGTTTATCA GAGTTTCAA GGTTCGCTCA CAGGTTTAC ATCCAGCAAG GAATCCGAGA CGTAAGGTG TGAGGTGTA CTACAACAT TTAAACGACAC
 492 E F Y D E L K W R L T E F T P R E K P E A N W C E L M I N Y F Q Q

2001 CACAGTTGGA TGGTTGCGGA AGTCTATAGT TTTGAGCCAA CATCTGGAT ACCTGGGCAC CTGTCAFACC ACTGTAAGGC ATTTGCCAT AAGTAATGAT
 CTGTCACCT ACCACGCT TCAGATATCA AAACCTCGTT GTAGACCTAA TGGACCCGTT GACAGTATGG TGACATTCG TAAACGGTA TTCATTTACTA
 459 P C N S P Q P L R Y N Q A L M Q I V Q A G T M G S Y P M K G Y T I I

2101 TTCATAAGA AGGATTCCAA ATGACCATAC ATCGGACTTA ATGCTGAAT TATTAATAGC AATGGCTTCG GCGCAGTCC ACTTCACCG CAGCTTTAT
 AAGTATTTCT TCCTAAGGTT TACTGATATG TAGCCTGAAT TACGACTTAA ATAATGATGC TTACCGAAGC CCGGTGTCAGG TGAAGTGCC GTCGAATTA
 425 E Y L L L I G F S W V D S K I S F K N S R I A E P A T W K V P L K I

2201 TCGTGTCTAG ATTATAGAT GTCTTCAATTA TCTACCTTAA AACTCTGGC AAGTCCAAA TCTGCTACTT TGTAGATATT ATGTTACCA ACAGGACAT
 AGCACAGATC TAAGTATCTA CAGAGTATAT AGATGGAAT TTGAGACCG TTCAGGTTT AGACGATGAA ACATCTATAA TACAAGTGT TGTCTCTGTA
 392 E H R S E Y I D E N D V K F V R A L G F D A V K Y I N H E G V L V

2301 TTCTGGCAGC CAGATCTCTG TGAATGATG TCCGAGACTC CAGATAGGCC ATTCCAGAG CAACCTGTGC CGCCATGCTT ACCTGTTGAG TCAGATGGAT
 AAGACCGTCG GTCTAGAGAC ACTTACATCA AGGCTCTGAG GTCTATCCG TTAGGTCTGC GTTGACACG GCGGTACAGA TGGACAATC AGTCTACCTA
 359 N R A A L D R H I Y N R S E L Y A M Q S A V Q A A M D V Q Q T L H I

FIGURE 5D

2401 TTTTGATCCA GTGTCATTTT GGAGATATTC TTGCAGACTT CCAATGCTCA TCAACTCTGT ANTAATATAA ATTGGATCTT CTAAGTGCA AACAGCATAA
 AAACTAGGT CACAGTAATA CCTCTATAG AACGTCTGAA GGTAACAGGT AGTTGAGACA TTATTATATT TAACCTAGAA GATTCACGT TTGTCGTATT
 325 K S G T D N Q L Y E Q L S G H R M L E T I I Y I P D E L T C V A Y

2501 AGCTGGATAA GCTTTGGATG TCTTAGGTTT TTCATTATCT GTGCTCCCT CAGGAAGTCA TTTGGATCCA TTGAACCTGG TTTTAATGTT TTCACTGCTA
 TCGACCTATT CGAAACCTAC AGAATCCNAG AAGTAATAGA CACGGAGGGA GTCTTCACT AACTTAGGT AACTTGACC AAAATTACAA AAGTGACGNT
 292 L Q I L K P H R L N K M I Q A E R L F D N P D M S G P K L T K V A

2601 CTGGAGTGGT ATTGTTCCAC AGACCTTCCC ATACTCGCC AACTGACCA GATCCCAATC GCTTCAGNAG CTGTATGGAG TTGCGGTCTA TCTCCCATTTG
 GACCTCACCA TAACAAGGTG TCTGGAAAGG TATGAGCGG TTTGACTGGT CTAGGTTAG CGAAGTCTTC GACATACCTC AACGCCAGAT AGAGGGTAAC
 259 V P T T N N W L G E W V E G F Q G S G L R K L L Q I S N R D I E W Q

2701 GTCCACGGTT TTATACGACA ATCAAAATGG AGCTGGGACC TGGATCTTTA AGCATGGTTT CCCCAGCTTG ACACACAGGC CGTCACTTGT CTTGGTGTAG
 CAGGTGCCAA AATATGCTGT TTAGTTTACC TCGACCTGG ACCTAGAAAT TCGTACCAAA GGGTTCGAAC TGTGTGTCG GCAGTGAACA GAACCAATC
 225 D V T K Y S L D F P A P V Q I K L C P K G L K V C L G D S T K T Y

2801 TGGCTCACAA ATTGTTTCAG TGTGAAAG ATTCTCTTTC GGTGAGAAA AAATCCCTCT TCATCCAGTC TTTTAATTT GTAGTGTGTT ACAACTGCTC
 ACCGAGTGGT TAAGCAAGTC ACACTTTTTC TAAGNAGNAG CGCACTCTTT TTTAGGGGGA AGTAGGTGAG AAAATTAGA CATCACAAA TGTGTGACGAG
 192 H S V F E N L T S F I R R R T L F F G G E D L R K I R Y H K V V A

2901 CATCTAAAC TGAAGAGAG AATCTCTCTT TTTGGCTTTC ACTTCTCTG ATTAGAAAG AACCGGTCTT GTTCTCTGAA TATAATAGTT GTTCTCTGCG
 GTAGATTTTG ACTTCTCTC TTAAAGAGGA AAACCGNAG TGAAGAGAG TANTCTTTC TTGGCCAGAA CAAAGACTT ATATTATCAA CAAAGAGACG
 159 G D L V S L S F E G K Q S E S E R I L F S G T K N E S Y L L Q K E A

3001 ATCTGATCTT CCGATTGCTC CAAGAACA CGGCTCTGCC TGTAGGCTTC TGTCTCAGC CACGTAGTTA GAAGGATAT AGCCTGTAG TTGCTGACTG
 TAGACTAGNA GGCTAACGAG GTTCTCTTGGT GCCGAGACGG ACNTCCGAG ACAGAGTGG GTGCATCAAT CTCTCTTATA TCGGAACATC AACGACTGAC
 125 D S R G I A G F F W P E A Q L S R D E A V Y N S P I Y G Q L Q Q S

3101 GAGCCATCTC GTCTTTTCTC CAAGTGTCTG GCAAAACCACC AGCCTCATG CAAAGTGTCC AGAAGTGA GTTTGTACC TGTCTGGAG CTCAAGTCTT
 CTCGGTAGAG CAGAAAGAG GTTCACAGAC CGTTTGGTGG TCGGGAGTAC GTTTCACAGG TCTTGAACCT CAAACAGTGG ACGAGCCTTC GAGTTCAGGA
 92 S G D R R K E L H R A F W W G E H L T D L V Q L K D G A R F S L D

FIGURE 5E

3201 CAGCAGTCCG AGCTGGTAA TCACAAGA CCACAAGTA GTGGCCATGC CTCTGTGACT GGGGAGAGCA AAGGGCCCT GATTTTCAA TCACGGTTGA
 GTCGTCAGGC TCGGACCAAT AGTTGTTC A V F Y H G H R Q S Q P S C L A O P N E I V T S
 59 E A T R A Q Y D P L A

3301 CTGTCTGCC TCCGTGDACA AACAGGGAG ATAGGTTCT AGTACTCCC AGAGCCTCTG ACAGATGTTG CTCATTTGCG CTGTGGGG AGAAGAGGAG
 GAACAGACGG AGGCACCTGT TTGTCCCTC TATCCCAAGA TCCATGAGGG TCTCGGAGAC TGCTACAAAC GATTAACACG GNAACCACCCC TCTTCTCCTC
 25 K D A E T S L C P L Y P E L Y E W L R Q C I N S M

3401 CAGGGTTCT CCTCTCCC TTAGTCTCTG CGATCCACCT TATCTTCTT TATCTTCTT CACCAGGCAA CTTTGAAGTC AGCACCNACT CACCATACCT CGAGAGGTAT
 GTCCCGNAGA GGGAGAGGG AATCAGAGAC GCTAGGTGGA ATAGAAGGA GTGTCCGTT GAACTTCAG TCGTGGTTGA GTGTATGAA GCCTCTCAT

3501 GCAAGTCCC GTTTCAGATC AGTCCAGCAG CTGGGTGCA GCAAGTCTA CCTGGAGAGA CTTACCGGT TGCTTCTGT GGTGGAGGT GCTACCCCGA
 CGTTTCAGG CAAAGTCTAG TCAGGTGTC GACCCACGT GTTCAAGAT GAGCTCTCT GAATGGCCGA ACGAAGAGCA CCGACCTCCA CGATGGGGT

3601 GGCAAACTG AGCAGAGCT GGGAGCTGC TCACTAGGA GGTGCTTTT CTCTTATCT GCTTAAGAT CCACAACAA AAATAAATA AATTAAAG
 CCGTTTGGAC TCGTCTCGA CCGTCTCGA AGTGATCTT CCACAGAAA GNAGAATAGA CGAATTTCTT GGTGTGTTT TTTATTTTAT TTTAATTTTC

3701 GGCTTTATTT AGACAAATAT CTGAGAACAG AATGGTGCCA TCTTGCCCTT TGTCCCAATA AAAAGTTAGC AAGAGGAAGC TACTAACCCC TGGTAAACC
 CCGAAATAAA TCTGTTTATA GACTCTTGT TTAACACGGT AGACGGAA ACAGGTTAT TTTTCAATCG TTCTCTTCTG ATGATTTGGG ACCATTTTGG

3801 TCCACGTCTT GCTTTGCGCA GGTGAGTCT GAGGATCTT CCATACCTAC CAGTCTGCG CCTGCAGTTC GCGGCCGCGA CTCTAGAGTC GACCTGCAGA
 AGGTGCAGAA CGAAGCGGT CCCAGCTGAG CTCCCTAGNA GGTATGGATG GTCNAGACGC GACGTCCAG CCGCCGCGCT GAGATCTCAG CTGGACGTCT

3901 AGCTTGGCG CCATGGCCCA ACTGTTTAT TGCAGCTTAT AATGGTTACA AATAAGCAA TAGCATCACA AATTCACAA ATAAAGCATT TTTTTCACATG
 TCGAACCGC GGTACCGGT TGMCAATA ACGTGCAATA TTACCAATGT TTATTTCGT TTAAGTGT TATTTCGTAA AAAAGTGAC

19/35

FIGURE 5F

4001 CATTCCTAGTT GTGGTTTGTG CAAACTCATC AATGTATCTT ATCATGTCTG GATCGGGAAAT TAATTCGGCG CAGCACCATG GCCTGAAATA ACCTCTGAAA
 GTAAGATCAA CACCAACAG GTTTGAGTAG TTACATAGNA TAGTACAGAC CTAGCCCTTA ATTAAGCCGC GTCTGTGTAC CGGACTTTAT TGGAGACTTT

4101 GAGGAACTTG GTTAGGTACC TTCTGAGGCG GAAAGAACCA GCTGTGGAAT GTGTGTCAAT TAGGGTGTG AAGTCCCCCA GGCTCCCCAG CAGGCAGAAG
 CTCCTTGAAC CATCCATGG AAGACTCCGC CTTTCTTGGT CGACACCTTA CACACAGTCA ATCCACACC TTTCAAGGAT CCGAGGGGTC GTCCGTCTTC

4201 TATGCAAGC ATGCATCTCA ATTAGTCAGC AACCAAGTGT GGAAGTCCC CAGGCTCCC AGCAGGCAGA AGTATGCATA GCATGCATCT CAATTAGTCA
 ATACGTTTC TACGTAGAT TAATCAGTC TTGGTCCACA CCTTTCAGGG GTCCGAGGG TCGTCCGTCT TCATACGTTT CGTACGTAGA GTTAATCAGT

4301 GCAACCATAG TCCCGCCCT AACTCCGCC ATCCCGGCC TAACTCCGCC CAGTTCGCC CATTCGCCG CCCATGGCTG ACTAATTTT TTTATTTATG
 CGTTGGTATC AGGCGGGGA TTGAGGCGG TAGGGCGGG ATTGAGCGG GTCAAGCGG GTAAGAGCG GGTACCGAC TGAATTAATAA AATAAATATC

4401 CAGAGGCCGA GGCGGCTCG GCCTCTGAGC TATTCAGAA GTAGTGAGGA GGCTTTTGT GAGGCTTAG GCTTTGCAAA AAGCTGTTAA CAGCTTGGCA
 GTCTCCGCT CCGCGCGAGC CGGAGACTCG ATAAGTCTT CATCACTCTT CCGAATAAAC CTCGGATCC GAAAACGTTT TTCGACNAIT GTCGMACCGT

4501 CTGGCCGTG TTTTACAACG TCGTGACTG GAAACCCCTG GCGTTACCCA ACTTAATCGC CTTCAGCAGC ATCCCCCTTT CGCCAGCTG CGTAAATAGCG
 GACCGGCAGC AATATGTTGC AGCACTGACC CTTTGGGAC CGCAATGGGT TGAATTAGCG GAACGTCTG TAGGGGAAA GCGGTCCACC GCATTTATCG

4601 AAGAGGCCCG CACCGATCG CCTTCCCAAC AGTTGCGCAG CCTGAATGCG GAATGGCGC TGAATGGTA TTTCTCTCTT ACGCATCTGT GCGGTATTTT
 TTCTCCGGC GTGGCTAGCG GGAAGGTTG TCAACGCTC GJACTTACC CTTACCGCG ACTACGCCAT AAAAGAGGAA TCGGTAGACA CGCCATAAAG

4701 ACACCGGATA CGTCAAGCA ACCATAGTAC GCGCCCTGTA GCGGCGCATT AAGCGCGCG GGTGTGTG TTACGCGCAG CGTGACCGCT ACCTTGTCCA
 TGTGGGTAT GCAGTTTCTGT TGGTATCATG CCGGGACAT CGCCCGGTAA TTCGCGCCG CCACACCACC AATCGCGCTC GCACGTGGCA TGTGAACGGT

20/35

FIGURE 5G

4801 GCGCCCTAGC GCGCGCTCCT TTGCGTTTCT TCCTTTCCTT TCTCGCCACG TTGCGCGGCT TTCCCGGTCA AGCTCTAAAT CGGGGCTCC CTTTAGGGTT
CGCGGGATCG CGGGCGAGGA AAGCGAAGA AGGNAAGAA AGAGCGGTGC AAGCGGCCGA AAGGCGCAGT TCGAGATTTA GCCCCCGAGG GAAATCCCAA

4901 CCGATTTAGT GCTTTACGGC ACCTCGACCC CAAANAACCT GATTGGGTG ATGTTTCACG TAGTGGGCCA TCGCCCTGAT AGACGGTTTT TCGCCCTTTG
GGCTAAATCA CGAATGCGG TGGAGCTGG GTTTTGTGAA CTAACCCAC TACCAGTGC ATCACCCTGGT AGCGGACTA TCTGCCNAAA AGCGGGAAC

5001 ACGTTGGAGT CCACGTTCTT TAATAGTGA CTCTTGTTC AAATGGAAC AACACTCAAC CCTATCTCG GCTATCTTT TGAATTAATA GGAATTTG
TGCAACCTCA GGTGCAAGAA ATTATCACCT GAGACNAGG TTTGACCTTG GTGTGAGTTG GATAGAGCC CGATAAGAA ACTAANTATT CCTAANAACG

5101 CGATTTCCGC CTATTGGTTA AAAATGAGC TGATTTAACA AAATTTAAC GCGAATTTTA ACAATATTT AACGTTTACA ATTTTATGTT GCACTCTCAG
GCTAAGCCG GATAACCAAT TTTTACTCG ACTAATTTGT TTTTAAATTT CGCTTAAAT TGTTTTATNA TGCANAATGT TAAAAATACCA CGTAGAGTC

5201 TACNATCTG TCTGATGCC CATAGTTAAG CCAGCCCCGA CACCGGCCAA CACCGCTGA CGCGCCTGA CGGCTTGTG TGTCTCCCG ATCGCTTAC
ATGTTAGAC AGACTACGC GTATCAATTC GGTGCGGCT GTGGCGGCT GTGGCGACT GCGCGGACT GCCGACAG ACAGAGGCC TAGCGAATG

5301 AGACNAGCTG TGACCGTCTC CGGAGCTGC ATGTGTCAGA GGTTTTCACC GTCATCACCG AAACGCGGA GACGAAGGG CCTCGTATA CGCTATTTT
TCTGTTGAC ACTGGCAGG GCCCTCGACG TACACAGTCT CCNAAAGTGG CAGTAGTGG TTTGCGCGCT CTGCTTTCCC GGAGCACTAT GCGGATATAA

5401 TATAGGTTAA TGTCAATGATA ATATGGTTT CTAGACGTC AGGTGGCACT TTTCGGGAA ATGTGCGCG AACCCCTATT TGTATTATTT TCTAATATA
ATATCCAAAT ACAGTACTAT TATTACCAAA GAATCTGCAG TCCACCGTGA AAAGCCCTT TACACGCGC TTGGGATAA ACAATATAA AGATTTATGT

5501 TTCAATATG TATCCGCTCA TGAGACAATA ACCCTGATA ATGCTTCAAT AATATTGAA AAGGAGAGT ATGAGTATTC AACATTTCCG TGTGCCCCT
AAGTTTATAC ATAGCGAGT ACTCTGTTAT TGGAGCTATT TACGAGTTA TTATACTTT TTCTTCTCA TACTCATAG TTGTAAGGC ACAGCGGAA

21/35

FIGURE 5H

5601 ATTCCCTTTT TTGCGGCATT TTGCTTTCTT GTTTTGTCTC ACCCAGAAAC GCTGGTGAAA GTAAAAGATG CTGAAGATCA GTTGGGTGCA CGAGTGGGTT
TAAAGGAAAA AACGCCGTAA AACGGAAAGB CAUAAACBAG TGGGTCTTTG CGACCACCTTT CATTTTCTAC GACTTCTAGT CAACCCACGT GCTCACCCNA

5701 ACATCGAACT GATCTCMAC AGCGGTAAGA TCCTTGAGAG TTTTCGCCCC GAAGAACGTT TTCCNATGAT GAGCACTTTT AAAGTTCTGC TATGTGGCGC
TGTAAGCTTA CCTAGAGTTG TCGCCATTCT AGGAACCTCT ANAGCGGG CTTCCTTGCA AAGTTACTA CTCGTGAAA TTTCNAGACG ATACACCGCG

5801 GGTATTATCC GGTATTGACG CCGGGCAAGA GCAACTCGGT CGCCGCATAC ACTATTCTCA GAATGACTTG GTTGAGTACT CACCAGTCAC AGAAAJAGCAT
CCATAATAGG GCATAACTGC GGGCCGTTCT CGTTGAGCCA GCGCGTATG TGATNAGAGT CTTACTGAA CAACTCATGA GTGGTCAGTG TCTTTTCGTA

5901 CTTAGCGATG GCATGACAGT AAGAGAAATTA TGCAGTGCTG CCATAACCAT GAGTGATNAC ACTGCGCCA ACTTACTTCT GACAACGATC GAGGAGACCGA
GAATGCCCTAC CGTACTGTCA TTCTCTTAAT ACGTCACGAC GGTATTGTA CTCACTATTG TGACGCCGT TGAATGAAGA CTGTTGCTAG CCTCCTGGCT

6001 AGGAGCTAAC CGCTTTTGTG CACAACATGG GGGATCATGT AACTCGCTT GATCGTTGG AACCGAGCT GAATGAAGCC ATACCAACG ACGAGCGTGA
TCTTCGATTG GCGAAAAAC GTGTTGTACC CCTAGTACA TTGAGCGGA CTAGCAACCC TTGCGCTCGA CTTACTTGGG TATGGTTGC TGCTCGCACT

6101 CACCACGATG CCTGTAGCAA TGGCAACAC GTTGGGCAAA CTATTAACTG GCGAACTACT TACTTAGCT TCCCGGCMAC AATTATAGA CTGATGAG
GTGGTGCTAC GGACATCGTT ACCGTTGTG CAGCGGTT GATNATTGAC CGTTGATGA ATGAGATCGA AGGCCCTTG TTAATTATCT GACCTAGCTC

6201 GCGGATAAAG TTGCAGGACC ACTTCTGCGC TCGGCCCTTC CGGCTGGCTG GTTATTGCT GATAAATCTG GAGCGGTGA GCGTGGTCT CCGGTATCA
CGCCTATTTC AACGTCTGTG TGAAGACGCG AGCCGGMG GCCGACCGAC CAAATAACGA CTATTAGAC CTCGGCCACT CGCACCCAGA GCGCCATAGT

6301 TTGCAGCACT GGGGCCAGAT GGTAAAGCCCT CCCGTATCGT AGTTATCTAC ACGACGGGA GTACGGCAAC TATGATGA CGAATAGAC AGATCGCTGA
AACGTGCTGA CCGCGGTCTA CCATTCGGA GGGCATAGCA TCAATAGATG TGCTGCCCT CAGTCCGTTG ATACCTACTT GCTTATCTG TCTAGCGACT

22/35

FIGURE 51

6401 GATAGGTGCC TCACTGATTA AGCATTTGGTA ACTGTGAGAC CAAGTTTACT CATATATACT TTAGATTGAT TAAAACTTC ATTTTAAAT TAAAAAGGATC
 CTATCCACGG AGTGACTAAT TCGTAACCAT TGACAGTCTG GTTCAATGA GTATATATGA AATCTAATA AATTTGAG TAAAAATTAA ATTTTCTTAG

6501 TAGGTGAAGA TCCTTTTGA TAACTCATG ACCAAATCC CTTAACTGA GTTTTCCTC CACTGAGCGT CAGACCCCGT AGAAAAGATC AAAGGATCTT
 ATCCACTTCT AGGAAAAACT ATTAGAGTAC TGTTTTTAGG GAATTGCCT CAAAAGCAAG GTGACTCGCA GTCTGGGGCA TCTTTTCTAG TTTCCTAGAA

6601 CTTGAGATCC TTTTCTCTG CGCGTAATCT GCTGCTTCCA AACAAAAA CCACCGTAC CAGCGTGCT TTGTTTCCG GATCAAGAGC TACCACTCT
 GAACCTCTAGG AAAAAAAGAC GCGCATTAGA CGACGAACGT TTGTTTTTTT GTTGGCATG GTGCCACCA AACAAACGGC CTAGTCTCG ATGTTGAGA

6701 TTTTCCGAAG GTAACCTGCT TCAGCAGAGC GCAGATACCA AATACTGTC TTCTAGTGA GCCGTAGTTA GGCACCACT TCAAGAACTC TGTAGCACCG
 AAAAGGCTTC CATTGACCGA AGTCGCTCTG CGTCTATGCT TTATGACMAG AAGATCAGAT CGGATCAAT CCGTGCTGA AGTCTTGA ACATCGTGGC

6801 CCTACATACC TCGCTCTGCT AATCCTGTTA CCAGTGGCTG CTGCCAGTGG CGATAAGTGG TGTTTACCG GGTGGAATC AAGACGATAG TTACCGGATA
 GGATGTATGG AGCGAGACGA TTAGGACAAT GGTCAACCGAC GACGCTCACC GCTATTACG ACAGATGGC CCAACCTGAG TTCTGCTATC AATGGCCTAT

6901 AGGCGCAGCG GTCGGGCTGA ACGGGGGTT CGTGACACA GCCAGCTTG GAGCGAACGA CCTACACCGA ACTGAGATAC CTACAGCGTG AGCTATGAGA
 TCCGCGTCCG CAGCCCGACT TGCCCCCCTA GCACGTGTGT CGGCTGAACT CTGCTTGT GTATGTGCAC GATGTGCTATG TCGATACTCT

7001 AAGCGCCACG CTTCCCGAAG GGAGAAAGGC GGACAGGTAT CCGGTAAGCG GCAGGGTCCG AACAGGAGAG CGCACGAGG AGCTTCCAGG GGGAAACGCC
 TTGCGGCTGC GAAGGGCTTC CCTCTTTCCG CCTGTCCATA GGCAATTCG CGTCCCAGCC TTGCTCTCTC GCGTGTCTCC TCGAAGGTCC CCTTTGCGG

7101 TGGTATCTTT ATAGTCTCTGT CGGTTTTCG CACTCTGAC TTGAGCGTGG ATTTTGTGA TGCTCTCAG GGGGGCGAG CCTATGAAA AACGCCAGCA
 ACCATAGAAA TATCAGGACA GCCCCAAGCG GTGAGACTG AACTGCGAGC TAAAAACACT ACGAGCAGTC CCCCCGCTC GGATACCTTT TTGCGGTCTG

23/35

FIGURE 5J

7201 ACGCCGCCCTT TTTACGGGTC CTGGCCCTTT GCTGGCCTTT TGCTCACATG TTCTTTCCCTG CGTTATCCCC TGATTCTGTG GATAACCGTA TTACCGCCTT
TGCGCCGGAA AATGCCAAG GACCGGAAA CGACCGAAA ACGAGTGTAC AGAAGGAC GCAATAGGG ACTAAGACAC CTATTGGCAT AATGGCGGA

7301 TGAGTGAGCT GATACCGCTC GCCGAGCCG AACGACCGAG CGCAGCGAGT CAGTGAGCGA GGAAGCGGAA GAGCGCCAA TACGGAACCC GCCTCTCCCC
ACTCACTCGA CTATGGCGAG CGCGGTCGAC TTGCTGGCTC GCCTGGCTCA GTCACTCGCT CCTTGGCCTT CTGCGCGTT ATGCGTTTGG CGAGAGGGG

7401 GCGCGTTGGC CGATTCAATTA ATGCAGCTGG CACGACAGGT TTCCCGACTG GAAAGCGGGC AGTAGCGCA ACGCAATTAA TGTGAGTTAG CTCACTCAAT
CGCGCAACCG GCTAAGTAAT TACGTCGACC GTGCTGTCCA AAGGCTGAC CTTTGGCCCG TCACTCGGT TCGTTAATT ACACTCAATC GAGTGAGTAA

7501 AGGCACCCCA GGCCTTACAC TTTATGCTTC CGGCTCGTAT GTTGTGTGGA ATTGTAGCG GATAACAATT TCACACAGGA AACAGCTATG ACATGATTAC
TCCGTGGGT CCGAATGTG AAATACGAAG GCCGAGCATA CACACACCT TAACACTCGC CTATTGTTAA AGTGTGCTT TGTGATAC TGTACTAATG

7601 GAATTAA
CTTAATT

24/35

1 GCGGCGCGCAG AGAAGCAGA GGATGGGGCT TAGCAGCTGG CAGAGCCAGG AGCGGGGAGG TAGCAGAAAG ACCACAAGTA CAAAGAAGTC CTGAACCTTT
CGCCGGCGGTC TCTTTCGTCT CTTACCCCGA ATCGTCGACC GTCTGGTCC TCGCCCTCC ATCGTCTTC TGGTGTTCAT GTTCTTCAG GACTTTGAAA

101 GGTTTTGCTG CTGCAGCCCA TTGAGAGTGA CGACATGGAG CACAAGACCC TGAAGATCAC CGACTTTGGC CTGGCCCCGAG AGTGGCACAA AACCAACAAA
CCAAAACGAC GACGTCGGGT AACCTCACT GCTGTACCTC GTGTTCTGG ACTTCTAGTG GCTGAACCG GACCGGGCTC TCACCGTGT TTGGTGTGTT

201 ATGAGTGCCG CXGGCACCTA CXCTGGATG GCTCCTGAGG TTATCAAGGC CTCACCTTC TCTAAGGGCA GTGACGTCTG GAGTTTGGG GTGCTGCTGT
TACTCACGGC G7CCGTGGAT G7GGACCTAC CGAGGACTCC AATAGTTCCG GAGGTGGAAG AGATTCCCGT CACTGCAGAC CTCAAAACCC CACGACGACA

301 GGGAACTGCT GACCGGGGAG XTGCCATACC GTGGCATTGA CTGCTTGCT GTGGCCTATG GCGTAGCTGT TAACAAGCTC ACACTGCCAT CCATCCACCT
CCCTTGACGA CTGGCCCCC 7ACGGTATGG CACCGTAAT GACGGAACGA CACCGGATAC CGCATCGACA ATTGTTCGAG TGTACGGTA GGTAGGTGGA

401 GGCC
CCGG

FIGURE 6

25/35

FIGURE 7A

1 ATGAGAGCGT TGGGCGGGA GGGGGCGAG CTGGGCGTGG TGGTGGTTTT TTGTTGAAAG AGATTGAGGA GTATTATAAA TCAAGATGTC GGTGTGATCA
TACTGTGGA ACCGCGGCT GCGCGGCTG GAGCGGGAG AGCAATAAAA AAGAGTTAC TATAAAAGCT GATAATGTT AGTTCTAGAG GAGACACTAGT

101 AGTGTGTTTT AATCAATCAT AAGAACAAAG ATTCAATCAAT GGGGAAGTGA TCAATATATG CCAATGGTATC AAGAGCTGTC GGTGTGCGTT
TCACACAAAA TTAGTTAGTA TTCTTTTATG TAATATAGTA GCGTTCAAT AATAGTATAG GTTACCATAG TCTTAGGAGC CTTCTGAGC GCACACGAA

201 GAGAGCCGAG AGCTCAGGGA CAGTGTAGGA AGCTGTGCT GTGGAAAGTGG ATGTATCTTC TTCCATCACA CTGCAAGTTC TGGTGGATGC CCCAGGGGAC
CTCTGGGTC TGAGTGCCT GTACATGCT TGACAGGGA CAGCTTCAGC TACATAGAG AGGTAGTGT GAGCTTCAG ACCAGCTAG GGTGCCCTTG

301 ATTTCCTGTC TCTGGGCTCT TAAGCAGAG TCCCTGAAT GGAAGCACA TTTGATTTA CAAACAGAG GAGTTGTTTC CATGTCAT TTGAAAATGA
TAAGGAGAG AGACCCAGAA ATTGGTGTGG AGGAGTTAA GGTGTGCTT AAACTAAAT GTTTGCTTC CTCACAGAG GTACCGTAA ACTTTTACT

401 CAGAAACCA AGCTGAGGA TACCTACTTT TTATTGAGG TGAAGCTAG AATTACAAA TATTGTTAC AATGAGTATA AGAAATACC TGTCTTACAC
GTCTTTGCT TGAACCTTT ATGATGAAA AATAAGTTC ACTTGATGG TTAATGTTT ATAACAAATG TCACTCATAT TCTTTATGG AGAAATGTC

501 ATTAAAGA GCTTACTTTA GAAAATGGA AAGCAGGAG GCGCTGCTCT GCAATCTGA GAGGCTTCCA GAGCGATCC TGGATGGCT GCTTTGGGAT
TAACTCTCT GGAATGAAT CTTTTACCT TTTGCTGCTG GGGACAGA GTATAGACT CTGCGAAGT CTCGCTAGG ACCTTACCA CGAAGCGTA

601 TCACAGGGGG AAGCTGTAA AGAAGAAAT CGAGCTGTTG TTAAAAAGA GGAAGAAGT CTTCATGAAT TATTGGGAC GCACATAGG TGGTGTGCCA
AGTGTGCCC TTGACATTT TCTCTCTTCA GTTGACAGC AATTTTCT GCTTTTTC TAAAGCTTA ATAAAGCTG CCGTATATCC AGGACAGGCT

701 GAAATGAAT GGGCAGGGA TGCACAGG TGTTCAGAA AGATTAAT CAAGCTCTG AGACCATTT GGCACAAATTA TTTCTTAAAG TAGGGAACC
CTTTATTTCA GCGTGCCTT AGTGTGCTG AGAATGTTA TGTAGATTTA GTTTCAGAG TGTGTTTAA GTGTGTTAAT AAAGAAATTC ATGCTGCTTG

FIGURE 7B

801 CTTATATGATTA AGGTTGCAAA GCTTTTCATGTT GAACCAATGGA TTTCGCTGCTCA CCTTGGAAAT AGAACAACAA GCACTCGAGS AGGCGAACATA CTTTGAAGATG
 GAATACCTAT TCCACGTTTC GACACAGTACA CTTCGTAGCT AAGCCGAGT GGACCCCTAA TCTTTTGTTC CBTGAGCTCC TCCGTTGAT GAAACTCTAC

901 ATTACCTATT CACCAAAACAG AACTATAGATA CCGATTCTCT TTCTTTTCTT ATCATTAATG SCACGAACCG ACACCGGATA CTACACTTGT TCCCTCTTCAA
 TCAATGATAA GTTGTCTGTC TTGATATCTAT GCTTAAGATA AACHAAACA TAGTAATGAC CTTCTTTTUC TGTGCGCTAT GATGTGAACH AGGAUAAGTT

1001 AGCATCCAG TCAATCAGCT TTGGTTACCA TCGTAGAAA GGGATTATTA AATGCTACCA ATTCAAGTGA AGATTATGAA ATTGACCAAT ATGAAGAGTT
 TCGTAGGCTC AGTTAGCTGA ACCCAATGCT AGCATCTTTT CCTAATAAT TTACGATGCT TAAGTTACTT TAACTGCTTA TACTTCTCAA

1101 TTCTTTTCT GTCAGTTTA AAGCCTACCC ACAATCAGA TGTACTTGGA CTTTCTCTCS AAAATCATT CTTTGTUAGC AAAAGGCTCT TGATAACGGA
 AACAAAAGA CAGTCCAAAT TCTGATGCG TCTTTACTCT ACATGAGCT GGAACAGAGC TTTTAGTAAA GGAACACTCG TTTTCCCAIA ACTATTGCT

1201 TACAGCATAT CCAAGTTTTC CAATCATAG CACCAACGAG GAGATATAT ATTCCATGCA GAATATGAT ATGCCCAAT TACCANAATG TTCAAGCTGT
 ATCTCTATA GCTTCANAAC GTTAGTATTC GTGCTGCTGTC CTCTATATA TAAGGTAGCT CTTTACTAC TACGGTTAA ATGCTTTTAC AGTGGGACA

1301 ATATAAGAG GAAACCTCAA GTCTCTGCGAG AAGCTTGGC AAGTCAAGCG TCTGTCTCT CCGATGCTAT CCGATTACCA TCTTGACCT GTTAHAAGTG
 TATAATCTTC CTTTGAUTT CAGGAGCTTC TTGHAAGCG TTGATCTGTC AGHCAAAHA GCTACCTAT GGTAAATGCT AGAACCTGGA CTTTCTTAC

1401 TTCAACAGS TCTCCCACT GCACAAAGA GATCACAAGA GAAATCTTGA ATAAAHAGC TAACAGAAAA GTTTTTHAGC AGTGGTGTTC GAGCAGTACT
 AAGTCTGCTC AGAGGCTTGA CBTGCTCTCT CTAGTCTCTT CTTCAAGCT TATCTTTCG ATTCTCTTTT CACAAACCTG TCACCCACAG CTCTCATGA

1501 CTAAACATGA GTGAAGCCAT AAAAGGTTTC CTGCTCAAGT GCTGTGCTAT CAATTCCCTT GGCACATCTT GTGAGACGAT CTTTTTAAAC TCTCCAGGCC
 GATTTTCTACT CACTTCGCTA TTCTCCCAAG GACCACTTCA CACACCTAT GTTAAGGGA CCGGTAGAA CACTCTGCTA GGAANAATTG AGAGTCCGG

27/35

FIGURE 7C

1601 CTTTCCCTTT CATCCAAAGAC AACATCTCAT TCTATGCCAAC AATTGCTTTT TTTCTCTCTCT TCAATTGCTG TTTAACCCCTG CTAATTTGTC ACAAGTACAA
 GCAAGGGAAA GTAGGTTCTG TTCTATGATTA ACATAGCTTTC TTAACCAACA ACATAGCTTTC TTAACCAACA AATTGGGAC GATTAACAG TGTTCATGTT

1701 AAAGCAATTT AGGTATGAAA GGCAGCTACA GATGCTACAG GTGACCGCAT CCTCAGATAA TCGACTACTC TACGTTUATT TCAAGGAATA TUAATATGAT
 TTTGCTTAAA TGCATACTTT GATTCGATTT CTATCTATTT CACTGCTCTA GGAGTCTATT ACTCATCAAG ATGCANCTAN AGTCTCTTAT ACTTATACTA

1801 CTCAAAATGG AGTTTCCAAG AGAAATTTATGAGTTTGGGA AGGTACTAGG ATCAGGTGCT TTGGGAAAAG TCAATGAAGG AACAGCTTAT GGAATTAAGCA
 GAGTTTACCC TCNAAAGGTC TCTTTTAAAT CTCNAAACCT TCCATGATCC TACTCCACGA AACCTTTTC ACTACTTGGG TTGTGGAATA CCTTAATCGT

1901 AAACAGGAGT CTCAATCCAG GTTACCGTCA AATGCTGAA AGAARAGCA GACAGCTCTG AAAGAGAGG ACTCATGTCA GAACTCAAGA TGAATGACCC
 TTGTGCTCA GAGTTAGGTC CAATGGCAGT TTATCCGACT TCTTTTTCGT CTGTGAGAC TTCTCTCTCG TCACTACAT CTGAGTTCT ACTACTGGGT

2001 TGTGGGAGG CACGAGATA TTGTGAACCT GCTGGGGGCG TGCACACTGT CAGGACCAAT TTACTTGATT TTGGAATACT GTTGTATGG TGAATCTCTG
 CGACCTCTG GTGCTCTTAT AACACTTGGG CGACCCCGCG ACCTGTGACA GTCTGTGTTA AATGAATTA AACCTTATGA CAACGATACC ACTGAGAGAG

2101 AACATCTAA GAGTAAAG AGAATAATTT CACAGGACTT GACAGAGAT TTTCAAGGAA CACAATTTCA GTTTTACCC CACTTTCCAA TCACATCCAA
 TTGATAGATT CTTCAATTTT TCTTTTAAA GTCTCTGAA CCTGCTGAA AAGTTCTT GTGTTAAGT CAAAAATGG GTGAAAGTT AGTGTAGGTT

2201 ATTCCAGCAT GCTGGTTCA AGAAGGTTG AGATACACCC GACTCGGAT CAATCTCAG GCTTCTATGG GAATTCATT CACTCTGAAG ATGAATTTGA
 TAACTCGTA CGGACCAAGT TCTCTTCAAG TCTATGTGG CTTGAGCTTA GTTTAGAGTC CGAAGTACC CTTAAGTAAA GTGAGACTTC TACTTAACT

2301 ATATGAAAAC CAAAAAGGC TGAAGGAGA GGAGGACTTG AATGTGCTTA AATGTGCTTA TCTTCTTTCG TTTGCTATC AATGTGCTTA AATGTGCTTA
 TATCTTTTG GTTTTTCG ACCTTCTCT CTTCTGAG TTACACGAAT GTAACTCT AGAAGAAAG AACCTATAG TTCAACGCTT TCTTACCTT

FIGURE 7D

2401 TTTCTGGAAATTTAAATCTGTGTTGTCACAGCA'GACCTGGCCG CACGGAACGT GCTTGTCCAC CACGGGAAAG TGTGTAAAGAT ATGTGACTTT GATTTGGCTC
AAAGACCTTA ATTTCAGCAC ACAATGTCT CTGGACCGC GTTCCTTGA CGAACAGTGG GTGCCCTTC ACCACTTCTA TACACTGAAA' CCFANCCGAG

2501 CAGATATCAT GAGTGATTCC AACTATGTTGTCAGGGCAA TUCCTGTGTC CCTGTAAAT GATGCGCC CCAAGCCCTG TTTGAAAGGCA TCTACACCAT
CTCTATGTG CTCACTAAGGTTGATACAACTGTCGTTACGGGCAAGCGGCTTACCTTTTA' CCTACCGGGG GCTTTCGGAC AACTTCCGT AGATGTGTTA

2601 TAAUAUTGATGTCTGTTCATATGGAATATTACTGTGGAAATCTTCTCAC TTTGTGTGAA TCTTTACCTT GGCATTCCGG TTGATGCTAA CTTCTACAAA
ATTCTCACTA CAGACCAGTA TACCTTATAA TACACCTT TAAACATG ACCACACTT AGGAATGGCA CCGTAAGGCC AACTACGATT GAAGATGTTT

2701 CTGATTCAAAATGUAATTA AATGGATCAG CCAATTTATG CTACAGAAAG AATATACATT ATAATGCAAT CCTGCTGGC TTTTGACTCA AGGAACGGC
GACTAAGTTT TACCTAAAT TTACCTAGTC GGTAAATAC CATGTCTCT TATATGTA TATTACGTTA GGACGACCGG AACTGAGT TCTTTTGGC

2801 CATCCCTCCCATAATTTGACTTCGTTTTTAGGATGTCACT'GGCAGATGCA GAAGAGCGA TGTATCAGAA TGTGATGGC CGTGTTCGG ATGTCTCTCA
CTAGGAAGGGATTAAACTGA AGCAAAATC CTACAGTGA CCGTCTACGT CTTCTTCCCT ACATAGTCTT ACACCTACCG GCACAAAGCC TTACAGGAGT

2901 CACCTACCAAACAGGGGAC CTTTCAGCAGAGAGATGGATTTGGGGCTAC TCTCTCGCA GGTCTCAGGTC GAAGATTCTGT AGAGAACAA TTTAGTTTTA
GTGGATGTTT TGTCCGCTG GAAGTCTG TCTCTACCTA AACCCGATG AGAGAGCGCT CCGAGTCCAG CTTCTAAGCA TCTCTTGT AAATCAAAAT

3001 AGGACTTCATCCCTCCACCTATCCCTAACA GGTGTGAT TACCAAAACA AGTTTAATTT CATCACTAAA AGAAATCTA TTATCAACTG CTGCTTCACC
TCTGTGAAGTA GGGAGGTGGA TAGGATTTCT CCGACATCTA ATGTTTTGT TCCAATTAAT CTAGTGATTT TCTTTTAGAT AATAGTTGAC GACGAAGTGG

3101 AGACTTTCTCTAGAGAGCG
TCTGAAGAAGTATCTCTC

29/35

FIGURE 8A

1 TCGGCTCCA CCGGCCAGG GAGATCAGA CTTGKXKGG CAGAGCCCC CCAACTCAG TTCCATCCT ACCGAGTGA GCGCGGCCA TGGAGTCCG
 1 AGTCGAGOT GGGCGGTCC CTCTAGTCT GACCCCCCC GTCGCGGG GATTGATGTC AGGCTAGGA TGGCTCACT CCGCGCGGT ACCTCGAGC
 M E L R
 101 GGTGCTGCTC TGTGGGCTT CGTTGGCCG AGCTTTGGA GAGACCTCC TGAACACAAA ATTGMAACT GCTGATCTGA AGTGGTGAAC ATTCCCTCAG
 CACGACGAG ACACCCGAA GCACCCGCG TCTGAACTT CTCTGGAGC ACTTGTTT TAACCTTTGA CGACTAGACT TCACCCACTG TACGGAGTGC
 5 V L L C W A S L A A A I E E T L L N T K L E T A D L K W V T P P Q
 201 GTGGACGGC AGTGGAGGA ACTGAGCGC CTGATGAGG AACAGCAGC CTTGTCACC TACGAGTGT GTGAGTCCA GGTGCCCCG GCGCAGGCC
 CACCTGCCG TCACCTCCT TGACTGCGG GACTACTCC TTGCTGTGTC GCACGCGTGG ATGCTTCA CACTGCAGT CCGACCGGGC CCGTCCCGG
 38 V D G Q W E B L S G I D E E G A V H V Y A T L R F T H L E C L S L P R A G
 301 ACTGCTTGG CACAGTTGG GTCCACGCG GRRGRCCT CCACGTTAC GCCAGCTGC GTTCACCAT GCTCGAGTGC CTGCTCCTGC CTGCGGCTGC
 TGACCGAGC GTGTCCAGC CAGGTTCCG CCGCGGCA GTTCACATG CGTGCAGC CGAGTGTGA CGAGCTCAGC GACAGGAGC GAGCCGAGC
 72 W L R T G W V P R R G A V H V Y A T L R F T H L E C L S L P R A G
 401 GCGCTCCTGC AAGGAGCCT TCACCTGCTT CTAATATGAG AGCGATGCG ACACGCGCAC GCGCTCAGC AGCGCTTGA TGGAGAACCT CTACATCAG
 CCGAGAGAG TTCTCTTGA AGTGGCAGAA GATGATATCT TCGCTACGCC TGTGCGGTG CCGGAGTGC GGTGCGACT ACCTCTTGG GATGTAGTTC
 105 R S C K E T P T V F Y Y E S D A D T A T A L T P A W H E N P Y I K
 501 GTGGACAGG TGGCGCGGA GCATCTCACC CGGAGCGCC CTGGGCGGA GCGCAGCGG AGGTGAAATG TCAAGAGCCT GGTCTCTGGA CCGCTCAGCA
 CACCTGTGC ACCGCGCT CGTAGAGTGG GCTTTCGCG GACCCGCT CCGGTGCCC TTCCACTTAC AGTCTTGA CCGAGAGCCT GCGAGTCTGT
 138 V D T V A A E H L T R K R P G A E A T G K V N V K T L R L G P L S K
 601 AGGTGCTCTT CTACTGCGC TTCCAGGACC AGGTGCGTG CATGGCGCTG CTATCCTGC ACCTCTTCTA CAAGAGTGC GCGCAGTGA CTGTGAACT
 TCCAGCCGA GATGAGCCG AAGTCTCTG TCCACGAC GTACCGGAC GATAGGAGC TGGAGAGAT GTTTTCAAG CCGGTCACT GACACTTGA
 172 A G F Y L A F Q D Q G A C H A L I S L H I P Y K K C A Q L T V N L

30/35

FIGURE 8B

801 TACTGCGGTG AGGATGGCCA GTGGCCGGA CAGCGGTCA CGGCTGCAG CTGTGCTCCG GGTTCGAGG CAGCTGAGGG GAGACCCAG TCGCGAGCTT
 ATGACGGCAC TCTTACCGGT CACCCGCTT GTTCTCCACT TCCGACGTC GACACGAGC CCCAGCTCC GTCACTCC CTCTGGTTC AGGCTCGGA
 218 Y C R E D G Q W A E Q P V T G C S C A P G P E A A E G N T K C R A C

901 GTGCCAGGG CACCTTCAG CCCCTTCAG GAGAGGTC CTCCAGCCA TCCGAGCCA ATAGCCACTC TACACCAAT GATCAGCG TCTGCCAGT
 CACGGTCCC GTGAGGTC GGGAGATC CTCTCCAG GACGGTGGT ACGGTGGT TATCGGTAG ATTGTGTAA CCTAGTGGC AGAGGTGAC
 272 A Q G T P K P L S G E G S C Q P C P A N S H S N T I G S A V C Q C

1001 CGCGTCGGG TACTTCGGG CAGGACAGA CCCCGRGT GCACCTGCA CCAGCCCTCC TTGGCTCCG CGAGCGTGG TTTCGGCTT GAGCGCTCC
 GCGCAGCCC ATGAGGCC GTGGTGTCT GGGGCCCA CTGGGACTT GTGGGAGG AGCCGAGGC GCCTGCACC AAGCGCGA CTTCGGAGG
 305 R V G Y F R A R T D P R G A P C T T P P S A P R S V V S R L N G S

1101 TCCCTGCACC TGGATGGAG TCGCCCGCTG GAGTCTGGT GCCGAGGA CTCACCTAC GCCTCCCTT GCGGAGTGG CGGACCCGA GCTCTCTTG
 AGGACGTGG ACCTTACCTC ACGGGGNC CTCAGACCAC CGCTCTCTT GGAGTGGATG CGGAGCGGA CGGCTGCAC GGTGGGCT CCGAGGAC
 338 S L H L E W S A P L E S G G R E D L T Y A L R C R E C R P G G S C A

1201 CGCCCTGCG GGGAGACCTG ACTTTGACC CGGCGCCCG GACCTGGTG GAGCCCTGG GAGCTAGCT AGGCTAGCT CCTTACTCTT
 GCGGACGCC CCTCTGAG TGAATGAG GATAGGAG GATAGGAG CCTGACAC CTCGAGCC ACCACAGC TCCGATGCA GACTGAGT GATATGGA
 372 P C G G D L T P D P G P R D L V R P W V V V R G L R P D P T Y T F

1301 TGAGTCACT GCATTGAG GGTATCTTC CTTAGCAG GGGCCGTC CATTTGAGC TGTCATGTC ACCACTGAC GAGAGGTACC TCTGCGAGT
 ACTCCAGTGA CGTACTTGC CCATAGGAG GATAGGAG CCGGCGAG GTAACTCG ACAGTTACG TCGTACTCG CTCTCCATGG AGGAGTCAC
 405 E V T A L N G V S S L A T G P V P F E P V N V T T D R E V P P A V

1401 TCTGACATCC GGTGAGCG GTCTCACC ATCAGTTGA GCCTGCTG TCGTCTGACT GCGGACCCA GTGGGCTGT GCTGACTAC GAGGTCAAT
 AGACTGTAGG CCCACTGCG CAGGAGTGG TCGTCACT TCGTCACT CCGACGGAC CCGACAGG GCGGTGGT CACCCGACA GAGCTGATG CTCCAGTTA
 438 S D I R V T R S S P S S L S L A W A V P R A P S G A V L D Y E V K Y

FIGURE 8C

1501 ACATGATGAA GGGGGCTG GGTCCAGCA GGTGGGTT CTGAAGAG CTGAAGAAC GGGGAGCT GCGGGGCTG AAGCGGGAG CCACTACTT
 TGTACTCTT CCGCGGCTC CCAAGGTCT GTTACTGTA GACTTCTG AGTCTTTGG CCGTCTGA GCGCGGCTC TTGCGGCTC GGTGATGTA
 472 II E K G A E G P S H V R F L K T S E N R A E L R G L K R G A S Y L

 1601 GGTGCAGTA CGGCGGCT CTGAGGCGG CTATGRCCT TTGCGCAG ACATCAG ACAGACCCA CTGATGTA GCGAGGCTG GCGGAGCAG
 CCACGTCCAT GCGCGGCGA GACTCGGCC GATTCGCG GAGCGGTC TTGTAGTGC GTCTGGTT GACTACTCT CCGTCCGAC CCGCTCTGC
 505 V Q V R A R S E A G Y G P F G Q E H I S Q T Q L D E S E G W R E Q

 1701 CTGGCCCTGA TTGCGGCAC GGCAGTCGT GGTGCTCC GGTGCTCT TGTCTCTGT GGTCTATTGT GTGCACTTC TGTGCTCAG GAGCAGAGC ATGTGTAG
 GACCGGACT AACCGGCTG CCGTCAGCAC CCACACGAG ACACAGCCA CAGTACAC CAGGTCAAG AGCGAGTCT CTCCTCTC TTACCTCTC
 538 L A L I A G T A V V G V V L Y L V V L V V A V L C L R K Q S N G R E

 1801 AAGCAGATA TTGGACAA CAGGACAGT ATCTATCG ACATGACT AAGTCTACA TCGACCCCTT CACTATGAA GACCTAATG AGCTGTGAG
 TTGCTCTTAT AAGCTGTT GTGCTGTCA TAGAGTAGC TTGACATGA TTCCAGTGT ACCTGGGAA GTGATACTT CTGGATATC TCGACACTC
 572 A E Y S D K H G Q Y L I G H G T K V Y I D P F T Y E D P N E A V R

 1901 GGAATTTGCA AAGAGATCG ATGTCTCTA GGTCAAGAT GAAGAGTGA TTGTGCGAG TGAAGTTGCG GAGGTGTC GCGGCGGCT CAGGCGGCA
 CCTAAAGCT TTCTCTAGC TACAGAGGAT GCAATCTTA CTCTCCACT AACCATCC ACTCAACCG CTCACACAG CCGCGGCGA GTTCCGCGT
 605 E F A K E I D V S Y V K I E E V I G A G E F Q E V C R G R L K A P

 2001 GGAAGAAAG AGAGCTGTGT GGCATCAAG ACCGTGAAG GTTCTACAC GAGCGGAG CGGCTGAGT TTCTGAGCA GCGCTCCTC ATGCGGCACT
 CCTTCTCTC TTCTGACACA CGTTAGTTC TGGACTTCC CACCATGTC CTTGCGCTC GCGCACTCA AGACTCGCT CCGAGGTAG TACCGGCTCA
 638 G K K E S C V A I K T L K G G Y T E R Q R R E F L S E A S I M G Q P

 2101 TCGAGCACCC CAATATCATC CCGCTGAGG GGTGTGAC CACAGCAT CCGCTCATGA TTCTCAGAG GTTCAATGAG AACCGGCGC TGGACTCTT
 AGCTCTAGG GTTATAGTAG GCGGACTTCC CCGACAGTG GTTGTCTAC GGCAGTACT AGAGTGTCT CAATACTCT TGGCGGCGG ACCTGAGGA
 672 F H P N I I R L E G V V T N S H P V W I L T E F M E N G A L D S F

FIGURE 8D

2201 CTTGCGGCTA AACGACGAG AGTTACAGT CATCAGCTC GTGGCATGC TGGGGGAT CCGCTGGGC ATCGGTACC TTGGGAGAT GAGTACCTC
 GGAGCCGAT TTGCTGCTG TCAGTGCTA GTAGTCTGAG CACCGTAGG ACGCCCGTA GCGAGCCCG TAGCCATGG AACGCTCTA CTCGATGAG
 705 L R L N D G Q F T V I Q L V G M L R G I A S G M R Y L A E M S Y V

 2301 CACGAGACC TGGCTGCTG CAAATCCTA GTCAACATCA ACCTGCTG CAAATGTCT GACTTTGGC TTTCGGAT CCTGAGGAG AACCTCTCG
 GTGCTCTG ACCGAGAC GTTAGAT CAGTTTCGT TGGAGAGAC GTTTCACAGA CTGAACCGG AAAGGCTAA GAGCTCTC TTGAGAGGC
 738 H R D L A A R N I L V N S N L V C K V S D F G L S R F L E E N S S D

 2401 ATCCACCTA CACGAGCTC CTGGAGGAA AGATTCCAT CCGATGGACT GCGCGGAG CCAATGCTT CCGGAAGTTC ACTTCGGCA GTGATGCTG
 TAGATGGAT GTGCTGAGG GAGCTCTCT TCTAGGTA GGCTACCTGA CCGGCTCC GGTAAAGGAA GGCCTCAAG TGAAGCGGT CACTACGAG
 772 P T Y T S S L G Q K I P I R W T A P E A I A F R K F T S A S D A W

 2501 GAGTTACGG ATTGTGATG GAGAGTGT GTCATTTGG GAGAGCGGT ACTGGGACAT GAGCAATCAG GAGTGTATCA ATGCCATTA ACAGGACTAC
 CTCATGCCC TAACACTACA CCTCCACTA CAGTAACCCC CTCCTCGCA TCACCTGTA CTCGTAGTC CTGCACTAGT TAGGTAACT TGTCTGTATG
 805 S Y G I V H W E V M S F G E R P Y W D M S N Q D V I N A I E Q D Y

 2601 CCGCTGCCCC CCGCCGAGA CTGTCCACC TCCCTCCACC AGCTATGCT GACTGTGG CAGAAAGACC GGAATGCCG GCGCTGCTTC CCGGAGTGG
 GCGGACGGG GCGGGTCT GACAGGTGG AGGAGGTGG TCGATAGCA CCTGACACC GTCTTTCTGG CCTTACGGC CCGGCGAAG GGGTCCACC
 838 R L P P P P D C P T S L H Q L M L D C W Q K D R N A R P R F P Q V V

 2701 TCAGCGCCT GAGACAGATG ATCGGAAAC CCGCAGCCT CAAATCTG GCGCGGAGA ATGCGGGGC CTCACACCT CTCCTGACC AGCGGAGCC
 AGTCGCGGA CCTGTTCTAC TAGGCTTGG GCGGTGCGA GTTTAGCAC CCGGCTCT TAGCGCCCG GAGTGTGGA GAGGACCTG TCGCCGTCG
 872 S A L D K H I R N P A S L K I V A R E N G O A S H P L L D Q R Q P

 2801 TCACTACTA GCTTTTGGT CTGTGGGGA GTGGCTTGG GCCATCAAA TGGAGATA CGAAGAAAT TTGCGAGCG CTGGCTTGG CTCCTTGGAG
 AGTGATGAGT CGAAACCGA GACACCCCT CACCGAGCC CGGTAGTTT ACCCTCTAT GCTTCTTCA AGCGTGGC GACCGAACC GAGGAGCTC
 905 H Y S A F G S V G E W L R A I K M G R Y E S F A A A G F O S F E

FIGURE 8E

2901 CTGTGTCAGCC AGATCTCTGCTGACACCTG CTGCGAATCG GATTCACCTCT GCGCGGACAC CAGAGAAAT TCTTGGCCAG TGTCCAGCAC ATGAGTCCG
 GACCACTCGG TCTAGAGACG ACTCTGACG GAGGCTTACG CTGAGTGAGA CCGCCCTGTG GTCTCTCTTT AGAACCGGTC ACAGGTCTGT TACTTTCAGG
 938 L V S Q I S A E D I, L R I G V T L A G H Q K K I L A S V Q H M K S Q

3001 AGGCCAAGCC GCGAACCCCG GGTGCGACAG GAGGACCGC CCGCCAGTAC TGACCTCGAG GATCTCCCA CCGCAGGAG ACCGCTCTCC CATTTTCCCG
 TCGGTTCCG CCGTTGGGCG CCACCTGTG CTCTGGCG GCGCTCATG ACTGACGTC CTGAGCGGT GCGGTCCCTG TCGCGGAGCG GTAAAGGCC
 972 A K P G T P G G T G G P A P Q Y O P A G T P H P R D T A S P F S G

3101 GCGAGAGTGG GGACTCACAG AGGCCCCAG CCGTGTGCC CGCTGGATTG CACTTTGAGC CGTGGGTTG AGGAGTTGCG ATTTGAGAG GACAGGATTT
 CCGTCTACG CCGTGTGTC TCGGGGTC GCGACACGG GCGACCTAAC GTTAAACTCG GGCACCCAC TCCTCAACCG TTAAACCTCT CTGTCTAA
 1005 A E W G L T E A P S P V P R W I A L O A R G V R S W Q F G E T G F

3201 GGGGTTCTG CCATATATAG AGGGAAAT CACCCCCAG CCACCTCGG GAACTCCAGA CCAAGGTGA GGGGCTTT CCCTCAGGAC TGGGTGTAC
 CCGCCAGAC GGTATTATCC TCCCTTTTA GTGCGGTC GGTGAGGCC CTGAGTCT GTTCCACT CCGCGGAA GGGAGTCTG ACCCACACTG
 1038 G G S A I I G G E N H P P A T S G N S R P R V R A P F P Q D W V O P

3301 CAGAGGAAA GGAAGTCCC ACATCTCC AGCTCCCA GTGCCCCC TCACCTTGT GGTGCTTC CCGCAGACCA AAGAGAGTGT GACTCCCTTG
 GTCTCTTTT CTTACAGGG TTGTAGAGG TCGAGGGGT CACGAGGG AGTGAACTA CCGACCAAG GCGCTCTGT TTCCTCACA CTGAGGAGC
 1072 E E K E V P N I S Q P P Q V P P S P O W V R S R R P K R V O L P C

3401 CCAGTCCAG AGTGGGGGG CTGTCCACAG GCGCAAGAG GGTGTACG GCGGAGTAC AATCATTT GGGTTGTAG TCCCACTTG CTCTGTGAC
 GGTGAGGTC TCACCCCCC GACAGGGTCC CCGTTCTTC CCGACAGTCC CCGGTCACTG TTTTGTAC CCGCAACATC AGGTTGAGC GACGACAGT
 1105 Q L Q S G G A V P G G K K G C Q G P V T K S L G F V V P T C C H

3501 CACCAACTC AATCATTTT TTCCCTTGT AATGCCCTC CCGGAGTGC TGCCTCATA TTGAGGTTT TTGAGTTTG TTTTGCTCT TAATTTTCT
 GTGTTTGTG TTAGTAAAT AAGGAAAT TTACGGAGG GGGGTGACG ACGGAGTAT AACTCCAAA AACTCAAAAC AAAAACAGAG ATTAAAGA
 1138 H Q T Q S P F S L V N A P P P A A A A P I L K V F E F C F W S O F F S

FIGURE 8F

3601 CCGGTTCCC TTTTGTTC TCGTTTGT TTTCTACCG TCTTGTGAT AACTTGTGT TGGAGGAC CTGTTTCACT ATGCGCTGCT TTGCCCAAGT
 GGGCAAGGG AAAAAAAG AAGCAAAACA AAGATATGC AGGACAGTA TTGAACACA ACCTGCTTGT GACAAAGTGA TACCGAGGA AAGGTTCA
 1172 P F P P C F P V L F P V L F P Y R P C H N P V L E G T C P T H A S F A Q V

 3701 TTAACAGGG GCCATCATC ATGTCTTTT CAGAACAGT GCTTGTGCA TCCACATCC CCGACCCCG CCGGACCC CCAGCTGTG TCCTATGAAG
 ACTTGTCCC CGGTAGTAG TACAGACAA GGTCTTGTCA CGAACAGT AGGTGTAGG GGCCTGGGC GACCCCTGG GATTCCACAC AGGATACTTC
 1205 E T G A H H V C F Q N S A L V I P H P R T P P G T P K L C P H K

 3801 GGGTGTGGG TGAGGTAGT AAGGGGGG TAGTGTGTG TGAACCCAG AACGGACGC CGGTGCTGG AGGGTTCTT AATTATATT TAAAAAGTA
 CCCACACCC ACTCCATCAC TTTTCCGCC ATCAACCACC ACCTGTGTC TTGCTGTGG GCCACGACC TCCCAAGAA TTTATATAA ATTTTTCAT
 1238 G C G V R O O K G R O L V V E P R N G R R C L E G F L N Y I O K S N

 3901 ACTTTTGTG TAAATAAAG AATATGGAC GTGTCCGAC TCCAGGGTA AAAAAAAA AAAAAAAA
 TGAATACAT ATTTATTTG TTTTACCTG CACAGGTGG AGGTCCCAT TTTTTTTTT TTTTTTTTT
 1272 F L Y K O K K H G R V P A P G V K K K K K K

- FIGURE 9
ARNILVNSNLVCKVSDFGLSRFLEDDTSDPTYTSALGGKIPMRWTAPEAIQYRKFSAS
- FIGURE 10
NVLKSPNHVKITDFGLARLLEGEDEKEYNADGGKMPIKWMALCEIHRYRKFTTHQS
- FIGURE 11
NCHLAGDMTVCVADFGLSWKIYSGATIVRGCASKLPVKWLALGSLADNLYTVHS
- FIGURE 12
NCLVGKNYTIKIADFGMSRNLYSGDY
- FIGURE 13
TRNILVENENRVKIGDFGLTKVLPQDKKEYYKVKEPGESPIFWYAPESLTSLSFVSAD
- FIGURE 14
ARNILVNSNLVCKVSDFGMSRVLEDDPEAAAYTRGGKIPIRWTAPEAIYRKFTSASD

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 93/00586

I. CLASSIFICATION F SUBJECT MATTER (If several classification symbols apply, indicate all) ⁶		
According to International Patent Classification (IPC) or to both National Classification and IPC		
Int.Cl. 5 C12N15/12; C12N15/54; C12N9/12; //C12Q1/68, C12N15/11		
II. FIELDS SEARCHED		
Minimum Documentation Searched ⁷		
Classification System	Classification Symbols	
Int.Cl. 5	C12N ; C12Q ; C07K	
Documentation Searched other than Minimum Documentation to the extent that such Documents are included in the Fields Searched ⁸		
III. DOCUMENTS CONSIDERED TO BE RELEVANT⁹		
Category ¹⁰	Citation of Document, ¹¹ with indication, where appropriate, of the relevant passages ¹²	Relevant to Claim No. ¹³
A	<p>NEURON vol. 6, no. 5, May 1991, pages 691 - 704 LAI, C. & LEMKE, G. 'An extended family of protein-tyrosine kinase genes differentially expressed in the vertebrate nervous system' see the whole document</p> <p style="text-align: center;">---</p> <p style="text-align: center;">-/--</p>	1-7
<p>¹⁰ Special categories of cited documents : ¹⁰</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another claim or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> <p>"I" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.</p> <p>"A" document member of the same patent family</p>		
IV. CERTIFICATION		
Date of the Actual Completion of the International Search		Date of Mailing of this International Search Report
07 JUNE 1993		02 -07- 1993
International Searching Authority		Signature of Authorized Officer
EUR PEAN PATENT OFFICE		ANDRES S.M.

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)		
Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No.
A	<p>PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA vol. 88, no. 23, 1 December 1991, WASHINGTON US pages 10411 - 10415 HOLTRICH, U. ET AL. 'Two additional protein-tyrosine kinases expressed in human lung: fourth member of the fibroblast growth factor receptor family and an intracellular protein-tyrosine kinase'</p>	1-7, 13-15
A	<p>GENE vol. 110, no. 2, 15 January 1992, AMSTERDAM NL pages 205 - 211 BRAUNINGER, A. ET AL. 'Isolation and characterization of a human gene that encodes a new subclass of protein tyrosine kinases' see the whole document</p>	1-7, 13-15
A	<p>PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA vol. 87, November 1990, WASHINGTON US pages 8913 - 8917 PARTANEN, J. ET AL. 'Putative tyrosine kinases expressed in K-562 human leukemia cells' see especially clone JTK10</p>	1-7
A	<p>MOLECULAR AND CELLULAR BIOLOGY vol. 9, no. 4, April 1989, WASHINGTON US pages 1587 - 1593 HAO, Q.-L. ET AL. 'Isolation and sequence analysis of a novel human tyrosine kinase gene' see the whole document</p>	1-7, 16-18
P,A	<p>WO,A,9 214 748 (AMERICAN CYANAMID COMPANY) 3 September 1992 see the whole document</p>	1-12

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 93/00586

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
obscurities.
see additional page
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/210

OBSCURITIES

Claims 2,3,5 and 7 as filed in the application are in contradiction with the description. Indeed, they are dependent on claims 1,4 or 6 which are related to DNAs or proteins of human MEGAKARYOCYTIC origin, whereas they claim some sequences which are not from such an origin, if taking reference to page 8 of the description. Therefore the search of these claims has been made independently of claims 1,4 and 6.

Claims 14 and 15 relating to the nucleotide and the protein sequences of SAL-D4 are filed as being dependent on claim 10 which relates to SAL-S1. They have been interpreted and searched as depending on claim 13.

Claims 17 and 18 relating to the nucleotide and the protein sequences of LpTK 3 are filed as being dependent on claim 14 which relates to SAL-D4. They have been interpreted and searched as depending on claim 16.

Claims 20 and 21 relating to a DNA expression vector and a cell transformed therewith, are filed as being dependent on claim 17 which relates to a protein (LpTK 3). They have been searched as depending on claim 19.

Finally, claim 2, point a) relating to SAL-S1 refers erroneously to SEQ ID 7 which describes SAL-D4. It has been interpreted as referring on SEQ ID 5 and 17).

US 9300586
SA 69794

07/06/93

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO-A-9214748	03-09-92	EP-A- 0536350	14-04-93

СРОЧНО ПОЧТИ

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82